

ISSN: 2407-1501

PROCEEDING

INTERNATIONAL CONFERENCE ON EDUCATIONAL RESEARCH AND EVALUATION (ICERE)

“Assessment for Improving Students' Performance”

May 29 – 31 2016

Rectorate Hall and Graduate School
Yogyakarta State University
Indonesia





Organized by:
Study Program of Educational Research and Evaluation
Graduate School, Yogyakarta State University
in Cooperation with Indonesian Educational Evaluation Association (HEPI),
and Center for Educational Assessment (PUSPENDIK) Ministry of Education and Culture

Proceeding

International Conference on Educational Research and Evaluation (ICERE) 2016

Publishing Institute

Yogyakarta State University

Director of Publication

Prof. Djemari Mardapi, Ph.D.

Board of Reviewers

Prof. Djemari Mardapi, Ph.D.

Prof. Dr. Badrun Kartowagiran

Prof. Geoff Masters, Ph.D.

Prof. Frederick Leung, Ph.D.

Bahrul Hayat, Ph.D.

Jahja Umar, Ph.D.

Prof. Burhanuddin Tola, Ph.D

Bambang Suryadi, Ph.D

Editors

Ashadi, Ed.D.

Suhaini M. Saleh, M.A.

Titik Sudartinah, M.A.

Layout

Rohmat Purwoko, S.Kom.

Syarief Fajaruddin, S.Pd.

Address

Yogyakarta State University

ISSN: 2407-1501

@ 2016 Yogyakarta State University

All right reserved. No part of this publication may be reproduced without the prior written permission of Yogyakarta State University

<p>All articles in the proceeding of International Conference on Educational Research and Evaluation (ICERE) 2016 are not the official opinions and standings of editors. Contents and consequences resulted from the articles are sole responsibilities of individual writers.</p>

Foreword of the Chairman

Assalamualaikum wr. wb.

Good morning ladies and gentlemen.

Praise be to Allah who has given abundant blessings so that we can hold this international conference.

This conference is aimed at improving the quality of assessment implemented in schools and other institutions. The quality of assessment determines students' ways of learning, so that it is hoped that the quality of education improves. Besides, this conference is a means of information exchanges in the forms of seminars dealing with results of research in educational assessment and evaluation. The expectation is that there is always improvement in educational assessment and evaluation methods, including in it is the instrument – both cognitive and noncognitive instruments.

The participants of this conference are the lecturers and teachers who teach educational assessment and evaluation, practitioners of assessment and evaluation, and researchers of assessment and evaluation. This conference can be held in cooperation with the Graduate School, Yogyakarta State University, Association of Educational Evaluation of Indonesia (HEPI), and Centre for Educational Research, Ministry of Education and Culture of Indonesia, supported by the Australian Council for Educational Research (ACER), Intel, Intan Pariwara Publisher, and many other institutions. For this reason, on behalf of the Organizing Committee, I would like to thank the Rector of Yogyakarta State University, Prof. Dr. Rochmat Wahab, M.Pd., M.A., and the Director of Graduate School, Yogyakarta State University, Prof. Dr. Zuhdan Kun Prasetyo, M.Ed., and all other institutions for their assistance and contribution that have made this conference possible. I would like to thank HEPI's Local Coordination Unit and all sponsors for supporting this conference and also all the audience for participating in this conference.

To the committee members, both in Jakarta and Yogyakarta, I would like to thank them for the hard work they have performed and for the togetherness so that this conference can be held.

Last but not least, we apologize for all the inconveniences you might encounter during this conference. Please enjoy the conference.

Wassalamu'alaikum wr. wb.

Prof. Djemari Mardapi, Ph.D.

Foreword of the Chairman of Himpunan Evaluasi Pendidikan Indonesia (HEPI)

Assalamu'alaikum Wr. Wb.

Indonesian Association for Educational Evaluation (HEPI) is a professional organization in education holding in the high esteem the principles of professionalism and knowledge development in the field of educational and psychological measurement, assessment, and evaluation. HEPI was established in November 19, 2000 in Yogyakarta, with a vision to become a professional organization that excels in the field of evaluation and measurement in education and psychology in Indonesia. Its mission is to develop up-to-date methodologies of evaluation, assessment, measurement, and data analysis in education and psychology, as well as studies of policies and technical implementation of the field for improving Indonesian education quality.

As a professional organization, HEPI brings together experts, practitioners and interested persons in the field of evaluation, assessment, and measurement of education, psychology and other social sciences. HEPI is open to anyone who has the interest the field with no restriction in terms of educational background and working experiences. Hopefully, through HEPI, members of the association can sustainably develop themselves as professionals. The existence of HEPI is also expected to contribute to the improvement of the quality of national education through research, consultancy, seminar, conference, publication, and training for members of the organization and for public audiences.

HEPI organizes annual workshop and conference in cooperation with the Regional Chapter of HEPI and universities. In 2016, for the first time HEPI organized **International Conference on Educational Research and Evaluation: Assessment for Improving Student's Performance** in May 29-30 2016 in Yogyakarta. This conference is jointly organized by HEPI and Yogyakarta State University and supported by the Center for Educational Assessment the Ministry of Education and Culture, Australian Council for Educational Research (ACER), INTEL Indonesia, and Intan Pariwara Publisher.

It is important to note that the choice of the HEPI 2016 conference theme is driven by the fact that the quality of our national education is still under expectation as shown by the results from School National Exam and international surveys conducted by some international agencies. HEPI believes that a number of factors contribute to the low quality of national education, including low teacher's knowledge and skills in classroom and school assessment. Therefore, improving the competence of teachers in classroom and school assessment is urgently required. In this context HEPI as a professional organization and individual members of the organization have to play an active role in improving teachers' competence in quality learning assessment.

In line with 2016 conference theme, HEPI invited two respected guest speakers, namely, Professor Geofferey Masters, Ph.D., Director of the Australian Council for Educational Research (ACER), who presented a paper on Assessment to Improve Student Competency and Professor Frederick Leung, Ph.D., from the University of Hong Kong, who delivered a paper on the International Assessment for Improving Classroom Assessment.

As a tradition, in 2016 conference HEPI organized two pre-conference workshops. The first workshop is on the conceptual introduction of Rasch model by Jahja Umar, Ph.D., senior lecturer at the Faculty of Psychology, State Islamic University Jakarta and the second workshop was delivered by Heru Widiatmo, Ph.D., researcher at American College Testing (ACT) Iowa, United States on Measuring Higher Order Thinking Skills (HOTS).

On behalf of HEPI, I would like to express my heartfelt gratitude to Rector of the Yogyakarta State University, invited speakers, resource persons, HEPI regional chapters, sponsors, speakers, participants, invited guests, and organizing committee who have worked hard in making this international conference a success. Thank you very much for your participation and support and we are looking forward to seeing you in the next conference.

Last but not least, we hope that all of us get much benefit from this conference for enhancing Indonesian quality education through quality assessment.

Wassalamualaikum wr. wb.

Chairman,
BAHRUL HAYAT, Ph.D.

Table of Contents

Foreword of the Chairman	i
Foreword of the Chairman of Himpunan Evaluasi Pendidikan Indonesia (HEPI)	ii
Table of Contents	iii
Invited Speakers	
Assessment for Improving Student Performance <i>Prof. Geoff Master, Ph.D.,</i>	
International Assessment for Improving Classroom Assessment <i>Prof. Frederick Leung, Ph.D.</i>	
Educational Quality assurance For Improving Quality of Education <i>Bahrul Hayat, Ph.D.</i>	
Parallel Session Speakers	
I. Sub Themes:	
- Assessment Methods for Improving Student's Performance	
Assessment Model for Critical Thinking in Learning Global Warming Scientific Approach <i>Agus Suyatna, Undang Rosidin</i>	1
The Nationalism Attitude Assessment of Students of State Senior High School 1 Pakem Sleman <i>Aman</i>	8
The Design of Formative Assessment by Inquiry Based Learning in Improving Students' Self-Regulation <i>Asih Sulistia Ningrum, Chandra Ertikanto</i>	14
Exploring the Use of One Meeting Theme-Based Extended Response A Practical Critical Thinking Assessment Tool for Classroom Practices <i>Ayu Alif Nur Maharani Akbar, Rahmad Adi Wijaya</i>	20
Application of Instructional Model of Daily Assessment for Improvement of Processes Quality and Instructional Outcomes <i>Benidiktus Tanujaya</i>	25
Assessing Student's Pragmatics' Knowledge at Islamic University of Riau <i>Betty Sailun</i>	30
The Teacher's Performance in Learning Process Management And Chemistry Learning Difficulties Identification <i>Budi Utami, Sulistyo Saputro, Ashadi, Mohammad Masykuri, Nonoh Siti Aminah</i>	39

Components of Scientific Attitude for Teacher Observation in Physics Learning in Senior High School Elvin Yusliana Ekawati	43
The Development of Psychomotor Competency Assessment on Physics Education Student of Palangka Raya University Enny Wijayanti	48
Implementation of Authentic Assessment in Bahasa Indonesia Subject for Senior High School in West Sumbawa Eny Rusmaini	55
Summative Assessment Design through the PjBL to Improve Students' Higher-Order Thinking Skills Erlida Amnie	59
Assessment Model Multiple Intelligences Learning Approach in Primary School Mathematics Subjects Helmiah Suryani, Badrun Kartowagiran	67
Indicator Development of Learning Model Evaluation Instrument Herpratiwi, Tien Yulianti, Adil Fadlilah H, Bajawati	73
Performance Assessment in Model of Learning Superflex® Huriah Rachmah	77
The Identification of Teachers Difficulties in Implementing of 2013 Curriculum at Elementary Schools Ika Maryani, Sri Tutur Martaningsih	84
Aerobic Gymnastics, Fitness, and Academic Grade of Health Diploma Students from Remote Areas In Indonesia Lucky Herawati, Maryana, Suharyono	91
Analyzing the Authenticity of Authentic Assessment Luki Yunita, Salamah Agung, Eka Novi	97
Design of Performance Assessment Based on Problem Based Learning in Improving Students' Self Regulation Luthfi Riadina, Agus Suyatna, Undang Rosidin	100
Implementation of Performance Assessment to Increase Biology Learning Achievement by Using Inquiry Model Murni Sapta Sari	105
Teachers' Belief in Implementing Feedback for Students' Writing in ESP Classroom Nisrin Adelyna Darayani, Rini Amelia	111
Comparison of Character Value Between Lower Class and Upper Class at Salman Al Farisi 2 Elementary Integrated School Rosaria Ijranti, Farida Agus Setiawati	115
Authentic Assessment in the Learning of Social Studies Rudy Gunawan	122

The Implementation of Assessment Model Based on Character Building to Improve Discipline and Student's Achievement Rusijono	129
The Design of Performance Assessment Based Guided Inquiry for Empowering Students' Argumentation Skills Saiful Imam Ali Nurdin, Viyanti	136
The Influence of Class Climate and Self Concept towards Achievement Motivation and Physics Learning Result of Student at XI IPA Grade SMA Negeri 1 Kahu Satriani, Kaharuddin Arafah, Muris	142
Assessment Cognitive for Physic: Development of Misconception Physic Test for Junior High School in Bangka Barat with Polytomous Model (PCM) Sikto Widi Asta, Dedek Andrian	151
Identifying of Undergraduate's Analytical Ability about Electric Current in Transistor Using Isomorphic Assesment Sri Hartini, Dewi Dewantara, Misbah, Syubhan Annur	158
A Performance-Based Assessment as a Current Trend in ELT: Investigating Its Washback Effects on Secondary-School Student Learning Sumardi	162
Developing an Authentic Assessment Science Process Skills, Creative Thinking Skills and Manipulative Skills Supahar, Dadan Rosana, Zamzam F A, Ryani Andryani, Neviana Wijayanti	168
Using of Self Assessment to Determine Science Process Skill and Concept Attainment Through Inquiri Learning of 8th Grade Student on 21th Junior High School in Ambon Wa Nurlina, K. Esomar, I. H. Wenno	173
Development Evaluation Model and Technical Evaluation Management Program Mahad Aly in The College of Islamic Religious Affairs (PTKIN) Winarno	177
The Development of Vocational Interest Instrumen for Career Exploration of Junior High School Students Yudhi Satria Restu Artosandi, Sudji Munadi	182
Self-Assessment of Teachers of Mathematics Vocational High School in Yogyakarta City on the Performance Post-Certification Zuli Nuraeni	200

II. Sub Themes:

- The Use Of Psychometric Method for Majoring Student's Competence

The measurement Model of Historical Consciousness Aisiah	206
Anbuso: Practical Software to Perform Item Analysis Ali Muhson, Barkah Lestari, Supriyanto, Kiromim Baroroh	215
Estimating of Students Capability Growth in Vertical Equating with Rasch Model Test Anak Agung Purwa Antara	221

Diagnostic Test Characteristics of Learning Difficulties in Mathematics for Science Class 12th Grader Apri Triana, Heri Retnawati	225
Assessing Science Process Skills using Testlet Instrument Ari Syahidul Shidiq, Sri Yamtinah, Mohammad Masykuri	231
The Effect of Multiple Choice Scoring Methods and Risk Taking Attitude toward Chemistry Learning Outcomes (An Experiment at SMA Negeri 13 Kota Bekasi, West Java) Awaluddin Tjalla, Sari Fitriani	235
Development of Personal Integrity Scale: Construct Validity Bambang Suryadi, Yunita Faela Nisa, Nenang Tati Sumiati	242
Argument-based Validity of Situational Judgment Test for Assessing Teaching Aptitude Budi Manfaat	248
Horizontal Equating in Accounting Vocational Theory Test Based on Mean/Mean Method of Item Response Theory Dian Normalitasari Purnama, Sigit Santoso	253
The Effect of Number of Common Items on the Accuracy of Item Parameter Estimates with Fixed Parameter Calibration Method Dina Huriaty	259
Analysis of Inter-Rater Consistency in Assessment Final Project Fashion Study Program Emy Budiastuti	265
Using Fuzzy Logic to Select Item Test in Computerized Base Testing Haryanto	269
An Application of the Generalized Logistic Regression Method in Identifying DIF (Analysis of School Examination in Soppeng) Herwin	276
Effects of Complexity Matter and Grouping Students of the Statistics Analysis Capabilities Ismanto	284
Construct Validity of the TGMD-2 in 7–10-Year-Old Surakarta Children with Mild Mental Disorder Ismaryati	289
Measurement of the Quality of Mathematics Conceptual Understanding through Analysis of Cognitive Conflict with Intervention Iwan Setiawan HR, Ruslan, Asdar	296
Modification of Randomized Items Selection and Step-Size Based on Time Response Model to Reduce Item Exposure Level of Conventional Computerized Adaptive Testing Iwan Suhardi	302
Characterics of an Instrument of Vocational Interest Scales Kumaidi	310
Rasch Model Analysis for Problem Solving Instrument of Measurement and Vector Subject Mustika Wati, Yetti Supriyati, Gaguk Margono	315

Analysis of Mathematical Reasoning Ability of Elementary School Students Using Timss Test Design Noening Andrijati	320
The Accuracy of Testees' Ability Estimation of The Essay Test and Testlets in Mathematics Through The Graded Response Model (GRM) Application Purwo Susongko, Wikan Budi Utami	326
The Comparison of Logistics Model on Item Response Theory: 1 Parameter (1pl), 2 Parameters (2pl), And 3 Parameters (3pl) Rida Sarwiningsih, Heri Retnawati	333
Validity and reliability examination of indicators development materials instruction at Elementary School base on Curriculum 2013 Rochmiyati	342
Analisis Item Information Function on the Test of Mathematics Rukli	348
Misuses Cronbach Alpha On Achievement Tests Satrio Budi Wibowo	355
Item Discrimination of Two Tier Test on Hydrolysis of Salt Sri Yamtinah, Haryono, Sulisty Saputro, Bakti Mulyani, Suryadi BU	360
An Analysis of Test Quality by Using ITEMAN Tia Nur Istianah, Desrin Lebagi	366
An Analysis of Person Fit Using Rasch Model Yessica Mega Aprita, Yolanda Septiana	372
Detecting Students Learning Difficulties Using Diagnostic Cognitive Tests Yuli Prihatni	380

III. Sub Themes:

- Developing Instruments of Educational Assessment

Development and Implementation of Higher Order Thinking Skills Instruments in Physics Education A. Halim, Yusrizal	385
Developing Picture Series and Vocabulary to Increase English Speaking Skill Agustina Ellyana, Ketut Martini and Agus Risna Sari	390
Indonesian Adaptation Scale of Zung Self-Rating Anxiety Scale (SAS) Alfiannor Luthfi Hasain	394
Development Hypothetical Model Resources Management Studies Teachers of Hindu Religion Aris Biantoro, I Made Sutharjana, Wayan Sukarlinawati	399
Indonesian Adaptation of Organizational Commitment Questionnaire from Meyer & Allen, 2004 Baqiyatul Auladiyah	406

Creativity Problems Test Form Students Complete Description of Learning Connection with Learning Outcomes Counting Mathematics in Primary Darmiyati	411
Effectiveness Guided Discovery Approachment Through Cooperative Learning Think Pair Share (TPS) Type in Terms of Students' High Order Thinking Skill (HOTS) Deny Sutrisno	418
Indonesian Adaptation on Scale of Readiness for Organizational Change Dharan Atasya Rakhmat	421
Developing Achievement Tests in Physics For Classroom Assessment Dhien Astrini, Kumaidi	427
The Development of Evaluation Model Education Life Skill Program Out of School Education Edi Subarkah	434
Development of Performance Assessment in Guided Inquiry Learning to Improve Metacognitive Skills and Student's Achievement Endah Handayani, Sunarmi, Murni Saptasari	440
Design Student Development Work Sheet (Learning Cycle) 5E to Improve Student Learning Outcomes High School Class X Feryco Candra, Chandra Ertikanto	445
Development of Vocational Interest Scale: A preliminary study of the psychometrics properties* Firmanto Adi Nurcahyo	449
Contextual Approach Using Pictures as a Media Increased Result and Motivation of Mathematical Learning (Mathematical Learning of Fractional Addition by Equalizing the Denominator) Ihsana El Khuluqo, Ningrum Rosyidah	455
The Content Validity of the Evaluation Model in the Affective Domain in Islamic Education Instruments Iskandar Tsani	461
Developing Science Process Skill Instrument of Islamic Senior High Schools Kadir, Sri Wahyuningsih, Abd. Rahman A. Ghani	467
Online Exam Model of Item Response Theory Based Cat Using Moodle Learning Management System Khairawati	473
Developing an Accreditation Model of Secondary School Marjuki, Djemari Mardapi, Badrun Kartowagiran	483
Developing an Instrument for Assessing the Performance of High School Physics Teacher Nurul Fitriyah Sulaeman, Badrun Kartowagiran	490
Analysis Instruments Test Reading for Academic Purpose Students of English Education Unisnu Jepara Nusrotus Sa'idah, Hayu Dian Yulistianti	496

Learning Evaluation Model Design with Multiple Choice Tests for Field Studies Exact Sciences Nyenyep Sriwardani	502
Bhagavad Gita Video for Hinduism Education Lampung Nyoman Siti, I Komang Arteyasa, Ni Made Indrayani	506
Development of Authentic Assessment Instrument at Grade Four Elementary School in Malang Puri Selfi Cholifah, Muhardjito, Eddy Sutadj	511
Model Employee Performance Evaluation of Economics Graduate Degree in Bali Putri Anggreni	517
Hypothetical Model Development of Electrical Torso Learning Media Circulation System for Students Skill Formation of Critical Thinking and Scientific Attitude Senior High School in Lampung Timur Ririn Noviyanti, Sisca Puspita Sari Nasution	523
Developing a Creative Thinking Assessment Model for Kindergarten Teachers Risky Setiawan	531
Indonesian Adaptation Scale for Job Content Questionnaire (JCQ) Sandra Jati Purwantari	539
Development of Assessment Instruments of Art Painting Production Integrated With Character for Assessing Learners' Field Work Practice in Vocational High School Trie Hartiti Retnowati, Djemari Mardapi, Bambang Prihadi	546
Analyzing the Quality of English Test Items of Daily, Mid Semester and Final School Examinations in Bandar Lampung: (Assessment and Evaluation in Language Teaching) Ujang Suparman	556
Developing A Pedagogical Commitment Instrument Wasidi	567
Adaptation and Construct Validation of the Indonesian Version of the Utrecht Work Engagement Scale Yulia	574

IV. Sub Themes:

- Program Evaluation for Improving Quality of Education

The Effectiveness of The Boarding Teacher Professional Development Program: an Approach of Process Evaluation Friyatmi	579
The Effect of Formative Test Types and Attitudes toward Mathematics on Learning Outcomes Hari Setiadi, Sugiarto, Rini	584
An Evaluation Model of Character Education in Senior High School Hari Sugiharto, Djemari Mardapi	591

An Evaluation on the Implementation of Lesson Plans for Early Childhood Education Center (PAUD) Located Around IAIN Surakarta Hery Setiyatna	598
The Effect of Cooperative Learning Model Type Group Investigation with Self Assessment Reinforcement and Learning Interest toward the Physics Learning Result of Students at Grade Xi SMA Negeri 1 Watubangga Kolaka I Gede Purwana Edi Saputra, H.M. Sidin Al	602
Effect of Cognitive and Emotif Techniques in Counseling Rational Emotif Behavior Therapy toward Tendency Aggressive Behavior Based on Type of Personality Among Students of SMP Negeri 4 Denpasar I Wayan Susanta	611
THE EVALUATION OF THE SCHOLARSHIP DEGREE PROGRAM FOR THE ISLAMIC RELIGIOUS EDUCATIONAL TEACHERS AT SCHOOL Ju'subaidi	617
The Influence of Teacher Pedagogical Competence and Emotional Intelligence towards Motivation and Physics Learning Result of Student at XI IPA Grade SMA Negeri 1 Watansoppeng Kaharuddin Arafah, Adnani Yuni, Muris	624
Evaluating Policy Implementation Indicators in Decentralized Schools Lilik Sabdaningtyas, Budi Kadaryanto	633
Identification Critical Thinking Skills of SMA Muhammadiyah 1 Banjarmasin Students to the Matter Dynamic Electricity Misbah, Saiyidah Mahtari, Sayid Muhammad Hasan	641
The Influence of the Socio-Cultural-Based Learning Device to Student Academic Performance Muhammad Nur Wangid, Ali Mustadi	645
The Influence of Teacher Professional Competence and Interpersonal Intelligence Towards Motivation and Physics Learning Result of Student at XI MIA Grade Sma Negeri 1 Pangkajene Murniaty M, Kaharuddin Arafah, Subaer	651
Evaluation Study to Career Guidance Service-Program of Vocational High Schools in Banjarmasin Nina Permatasari, Djaali, Ma'ruf Akbar	660
Cipp Evaluation of The Learning in Cultural Dialogue During Unsoed Intercultural Summer-Camp Oscar Ndayizeye, Agrégé TEFL	666
Evaluating Basic English Test Items for Non-English Students from Teachers Perspectives Prihantoro	673
Is the German Language Text Too Short for the Senior High School Students? Ryan Nuansa Dirga, Primardiana Hermilia Wijayati	679

Evaluation of Managerial Leadership Ability of Senior High School Headmasters in Sleman Sabar Budi Raharjo, Lia Yuliana	686
Evaluation of Social Attitude Core Competence (KI-2) Implementation in State Elementary School in Yogyakarta Siti Aminah, Yulian Sari	691
The Evaluation of The Foreign Language Intensification Program for the Students of UIN Allauddin Makassar Sitti Mania	696
Evaluation of the Civilizing Moral Character Implementation in Elementary School Sulthoni	701
The Evaluation of 2013 Curriculum Implementation on Thematic Integrative toward Math Subject for Elementary School In East Lombok Syukrul Hamdi	706

ASSESSMENT MODEL FOR CRITICAL THINKING IN LEARNING GLOBAL WARMING SCIENTIFIC APPROACH

Agus Suyatna, Undang Rosidin

Physics Education Postgraduate Study Program, University of Lampung
asuyatna@yahoo.com

Abstract— Global warming may have an impact on food shortages, natural disasters, floods, economic disruption, massive migration, can even trigger a war. Learning and assessment of global warming is essential implemented to provide an understanding of the causes and impact of global warming on the life. Critical thinking assessment needed to foster the idea of solving the problem of global warming and its impact on life in the future. This study aims to: describe the students' understanding of the causes and effects of global warming for life, describing the need of critical thinking assessment, designing the critical thinking assessment on global warming learning using scientific approach that can foster the idea of solving the problem of global warming. Data were collected using a written test on the junior and senior high school students and a questionnaire to teachers of physics in Lampung province. The data analysis includes qualitative and quantitative analysis. Design assessment conducted through literature review and expert validation. The results showed: (1) the students' understanding of the causes and effects of global warming for life is still low, (2) teachers of physics require a model assessment of learning global warming using scientific approach to foster the idea of solving the problem of global warming, (3) critical thinking assessment on learning global warming using scientific approach, measure all indicators of critical thinking on every component of scientific approach using techniques and test forms adapted to the indicator.

Keywords: *assessment, critical thinking, global warming, scientific approach*

I. INTRODUCTION

Carbon emissions continue to rise will cause the thick of greenhouse gases in the atmosphere. This layer will forward the heat coming from the sun but will retain heat in the earth's surface is called the greenhouse effect. The greenhouse effect is very much needed by all living things on earth, because without it, the planet would be very cool. However, if the gases are excessive in the atmosphere, will lead to global warming (1). One of the most critical environmental subjects of 21st century is Global warming (2). Intergovernmental Panel on Climate Change (IPCC) has concluded that most of the increase in global average temperature is most likely caused by increased concentrations of greenhouse gases due to human activities. Global warming may have an impact on food shortages, natural disasters, floods, economic disruption, massive migration, can even trigger a war (3). Rising global temperatures are expected to cause other changes such as rising sea levels, increased intensity of extreme weather phenomena, as well as changes in the amount and pattern of precipitation. The effects of global warming include agricultural output, loss of glaciers and the extinction of various animal species (2). Greenhouse gases (water vapor, carbon dioxide, methane, etc) existing in the atmosphere, trapping sunrays and making earth's atmosphere warmer. Whereas some of these gases are naturally found in the atmosphere, human activity increases the amount of particular gases (4-6). There are two main approaches to slow the increase in greenhouse gases. First, protecting trees and planting more trees. Second, reduce the production of greenhouse gases, by reduce the burning of fossil. Both ways this can be done if there is an awareness of the community about the impact of global warming.

People's knowledge about global warming is still low. This is demonstrated by the behavior and activities of the people who contributed to the thick greenhouse gases. Most of these greenhouse gases produced by animal husbandry, burning of fossil fuels in motor vehicles,

modern factories, farms, Nitrogen Oxide (NO) from fertilizers, and the gases used in refrigerators and air conditioners (CFC), as well as plants electricity. The destruction of forests which should serve as storage of CO₂ is also a contributor to greenhouse gases because trees that die will release the CO₂ stored in their tissues into the atmosphere. Forest fires are also the largest contributor to greenhouse gases in Indonesia.

The public's understanding of the causes of global warming and its impact on life and the environment is very important. An understanding of this is expected to be able to form a disapproval of the behavior that can lead to global warming and foster the ideas to solving the impact of global warming. Therefore, people need to be educated about the causes of global warming and its impact on life and the environment. Learning communities can be started at high school students who are part of the community. As students are among the most trainable stratum of the society, it is essential to discover their knowledge and perception of greenhouse effect to prepare the best kind of educational program for them (7).

One of the educational efforts that can be done is applying critical thinking assessment on learning the scientific approach. Critical thinking assessment needed to foster the idea of solving the problem of global warming and its impact on life in the future. Critical thinking in very simply stated, the ability to analyze and evaluate information. Critical thinkers raise vital questions and problems, formulate them clearly, gather and assess relevant information, use abstract ideas, think open-mindedly, and communicate effectively with others. Critical thinking is an important and necessary skill because it can help you deal with mental and spiritual questions, and it can be used to evaluate people, policies, and institutions, thereby avoiding social problems (8). They are involved in dialog, debate, writing, and problem solving," as well as higher-order thinking, such as analysis, synthesis, and evaluation. The encouragement of critical thinking can be accomplished in any content area by modification of lectures and the incorporation of simple active learning techniques (9). Indicators of critical thinking skills are the skill of asking and answering questions, consider whether the source is reliable or not, identifying terms and consideration of a definition, determines the action to be taken, and the skills to make a conclusion (10). This indicator will be used to design assessment of global warming with a scientific approach

Assessments have various purposes such as assessment for learning, assessment of learning and assessment to learning. To assess the critical thinking skills needed authentic assessment. Authentic assessment is defined as a real assessment. Which measures the ability of students from all aspects. Authentic assessment is defined as an actual vote. It is necessary for assessment using a variety of techniques. A variety of assessment methods should be able to measure all aspects of a students knows and the students did. Authentic assessment based on scientific literacy is a form of assessment of real, meaningful for the students, are able to develop high-level thinking skills and scientific literacy containing dimensions (concepts, processes, and context) (11). Development of scientific literacy assessment instrument is based on an understanding of the concepts and methods of science, technology and the impact of science on the environment (12). The essence of this form of assessment is that the assessment is based on several sources. Some things that clearly characterize the application of authentic assessment is to measure all aspects of learning, including the process, performance and product (13). Benefits of using authentic assessment, as stated is as follows. First, the use of authentic assessment allows direct measurement of the performance of students. Second, authentic assessment give an opportunity to the students to construct learning results. Third, authentic assessment allows the integration of the activities of teaching, learning and assessment activities into one integrated package. Fourth, authentic assessment gives students the chance to show the results of their study, work performance, in a way that is considered the most (11).

Good planning should also be applied in the assessment activities that are part of the learning activities. Steps to be taken in the development of authentic assessment in global warming learning, which cover three aspects: (1) the assessment of the attitude that includes self-assessment techniques and journals; (2) an assessment of knowledge using written test with a form of multiple choice questions and essay; (3) The skills assessment includes an assessment of performance, project, and portfolio. Valuation techniques are selected according to indicators of critical thinking on global warming learning using scientific approach.

This study aims to: describe the students' understanding of the causes and effects of global warming for life, describing the need of critical thinking assessment, designing the critical thinking

assessment on global warming learning using scientific approach that can foster the idea of solving the problem of global warming impact.

II. RESEARCH METHOD

The sample consisted of teachers and students. Sample of students taken from 11 senior high school and 7 junior high school spread over the district in the Province of Lampung. Students who become the sample are students who have studied the topic of global warming and which have not. Details about the sample presented in Table 1. The total sample of teachers as many as 15 people become from 3 senior high school.

Table 1. THE BACKGROUND OF THE STUDENT SAMPLE

School and grade	Already studying global warming	N
Senior High School grade 11	yes	196
Senior High School grade 10	no	126
Junior High School grade 8	yes	165
Junior High School grade 7	no	86
Total		573

Data were collected using a written test about global warming knowledge in the form of an objective test consisting of 15 questions. Each question has three choices, true, false, and do not know. Before using, the instruments validity and reliability tested first. Data were analyzed using independent sample t-test, one way ANOVA, and multiple comparison.

Data collected from the teachers using questioner about how teachers teach global warming, teaching problem, and the need for teachers to global warming learning program. This data was analyzed qualitatively.

III. RESULTS AND DISCUSSION

A. Results

The results of the written test the understanding of global warming are presented in Table 2. Knowledge about global warming of student in Lampung Province was low. The average of score 3.6 of 10 scale. The average student only able to answer questions correctly as much as 36%. The highest average score obtained by the student of grade 11 who have learned about global warming.

Table 2. STUDENT SCORE OF GLOBAL WARMING KNOWLEDGE

School and level	Already studying global warming	N	average
Senior High School grade 11	yes	196	4.1
Senior High School grade 10	no	126	3.6
Junior High School grade 8	yes	165	3.5
Junior High School grade 7	no	86	3.2
Average			3.6

The learning experience of global warming influence have effect on test scores. This is indicated by one-way ANOVA test results, there was significant difference ($p = 0.000 < 0.05$) in scores obtained knowledge of global warming.

The results of a further test for the average difference in test results between grade and learning experiences of global warming are presented in Table 3. Average scores were significantly different that students grade 11 with grade 10, grade 8, and grade 7. otherwise there

was no significant difference. It shows that global warming learning at the senior high school make differences in learning outcomes, while at the junior high school level is not.

Table 3. . MULTIPLE COMPARISONS OF STUDENT

(I) Class	(J) Class	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2	1	-.4674 [*]	.16333	.023	-.8883	-.0465
	3	.1254	.15113	.840	-.2640	.5148
	4	.4194	.18502	.107	-.0573	.8961
1	2	.4674 [*]	.16333	.023	.0465	.8883
	3	.5928 [*]	.16923	.003	.1568	1.0289
	4	.8868 [*]	.20008	.000	.3713	1.4023
3	2	-.1254	.15113	.840	-.5148	.2640
	1	-.5928 [*]	.16923	.003	-1.0289	-.1568
	4	.2940	.19024	.411	-.1962	.7842
4	2	-.4194	.18502	.107	-.8961	.0573
	1	-.8868 [*]	.20008	.000	-1.4023	-.3713
	3	-.2940	.19024	.411	-.7842	.1962

Note:

*. The mean difference is significant at the 0.05 level.

1. Senior High School grade 11, already studying global warming
2. Senior High School grade 10, global warming has not learned
3. Junior High School grade 8, already studying global warming
4. Junior High School grade 7, global warming has not learned

Further testing is done to determine the effect of gender differences and interaction between the learning experience by gender using Two Way ANOVA. The test results showed no difference in learning outcomes due to differences in gender and no interaction between the learning experience of global warming by gender (Table 4).

Table 4. STUDENT TESTS OF BETWEEN-SUBJECTS EFFECTS

Dependent Variable: Score					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	49.572 ^a	7	7.082	3.461	.001
Intercept	6560.369	1	6560.369	3206.303	.000
Learning experience	37.582	3	12.527	6.123	.000
gender	.634	1	.634	.310	.578
Learning experience * gender	3.518	3	1.173	.573	.633
Error	1156.038	565	2.046		
Total	8848.530	573			
Corrected Total	1205.610	572			

a. R Squared = .041 (Adjusted R Squared = .029)

B. Discussion

From the first question, about global warming is a natural event that there are 68% of students answered that the events of global warming is not a natural event means that human intervention in it. Question number 2 and 3 are questions to dig information about gases triggers the greenhouse effect, 23% of students confirmed that hydrogen, helium and neon are greenhouse gases and 13% stated that animals such as cows produce greenhouse gases. It shows most students do not understand the gases triggers the greenhouse effect. Question 7 regarding ozone obtained information that 81% of students believe that the holes in the ozone layer of the earth will increase the greenhouse effect, this is a misconception of the greenhouse effect. Question number 14 regarding the result of the greenhouse effect, as many as 58% of students stated skin cancers are caused by the greenhouse effect, in fact, skin cancer is caused by depletion of the ozone layer.

Based on the results of tests analysis and interviews with students, it can be seen the following matters. Student understanding on global warming can be grouped into three part: causes, impacts and solutions of global warming. Among the three sections, most do not understand is the cause and determine the solution. They are generally difficult to link the causes, impacts and solutions.

Students generally know the cause of global warming associated with the increase in the average temperature of the Earth's surface caused by rising greenhouse gas emissions. Greenhouse gases cause a phenomenon called the greenhouse effect. In general, students can give the analogy of the greenhouse effect that they had experienced being in a closed car that is park under the scorching heat of the sun and feel the temperature on the car was very hot, but they were not able to explain exactly why it can happen.

Greenhouse gas emissions considered dangerous that should be controlled are: carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), O₃, CCL₂F₂, CCl₂F₂, and sulfur hexafluoride (SF₆). From the test results there are a variety of student responses regarding greenhouse gases are considered dangerous. From the students' responses show that students do not understand precisely gases triggers the greenhouse effect.

Students understand that the greenhouse effect will have an impact on global warming that will be followed by climate change, such as increased rainfall in some parts of the world, causing flood and erosion. Meanwhile, in the other hemisphere will dry season due to the temperature rise, but the students have not been able to explain exactly how the greenhouse effect affects climate change.

Most students concerned to existence of global warming. Various efforts are known by students to overcome the impact of global warming include reforestation and reducing air pollution. They do not know that simple act such as: walking or go to school by bike, turn off the lights after they finish studying, saving paper consumption, and others can reduce greenhouse gas contribution.

These findings were not different from the results of research to Iranian. The study showed high percentages of the students believed that if the greenhouse effect gets bigger "the Earth will get hotter", "there will be changes in the world's weather" and "polar ice will melt". These findings together with students' low level cognitive scores and the fact that students were not able to describe greenhouse effect correctly and completely indicate that students know some points about global warming but their knowledge is not deep-seated. Despite of assuming "holes in the ozone layer" as a cause of greenhouse effect by most of the students, near the same percentage of the students were aware of the fact that too much carbon dioxide and chlorofluorocarbons are causes of global warming. The students awareness about the causes of the greenhouse effect. While more than two third of the students knew "not using cars so much" and "planting more trees" are cures for global warming, about the same percentage of the students accepted this wrong idea that "using unleaded petrol is a way to reduce greenhouse effect", therefore students' knowledge about cures for greenhouse effect is not reliable too (7).

The results of the classroom observation indicate the learning process is centered on the teacher, the students tend to be passive in learning. The learning process tends to abstract, the teacher has not optimize the use of instructional media. Media used only pictures that are not attractive. Teachers have no implement a scientific approach in the learning process, as emphasized in the curriculum. From needs analysis questionnaire obtained information that the

teacher wanted a model of learning, assessment models, and various media to study global warming.

Based on the results of a literature review prepared an assessment models to foster critical thinking skills in understanding of global warming and looking for ideas to solving the impact of global warming as presented in Figure 1. There are five stages of global warming learning to cultivate critical thinking skills. The stages of learning global warming are as follow. Observer, to observed the phenomenon of global warming that there are around students. Questioning, asking questions about the causes of global warming phenomenon. Explore, conducting experiments to determine the relationship between the phenomenon of global warming and its causes. Infere, determine the factors that cause global warming and looking for ideas to reduce greenhouse gases. Communicate, convey ideas and attitudes toward behavior that causes global warming.

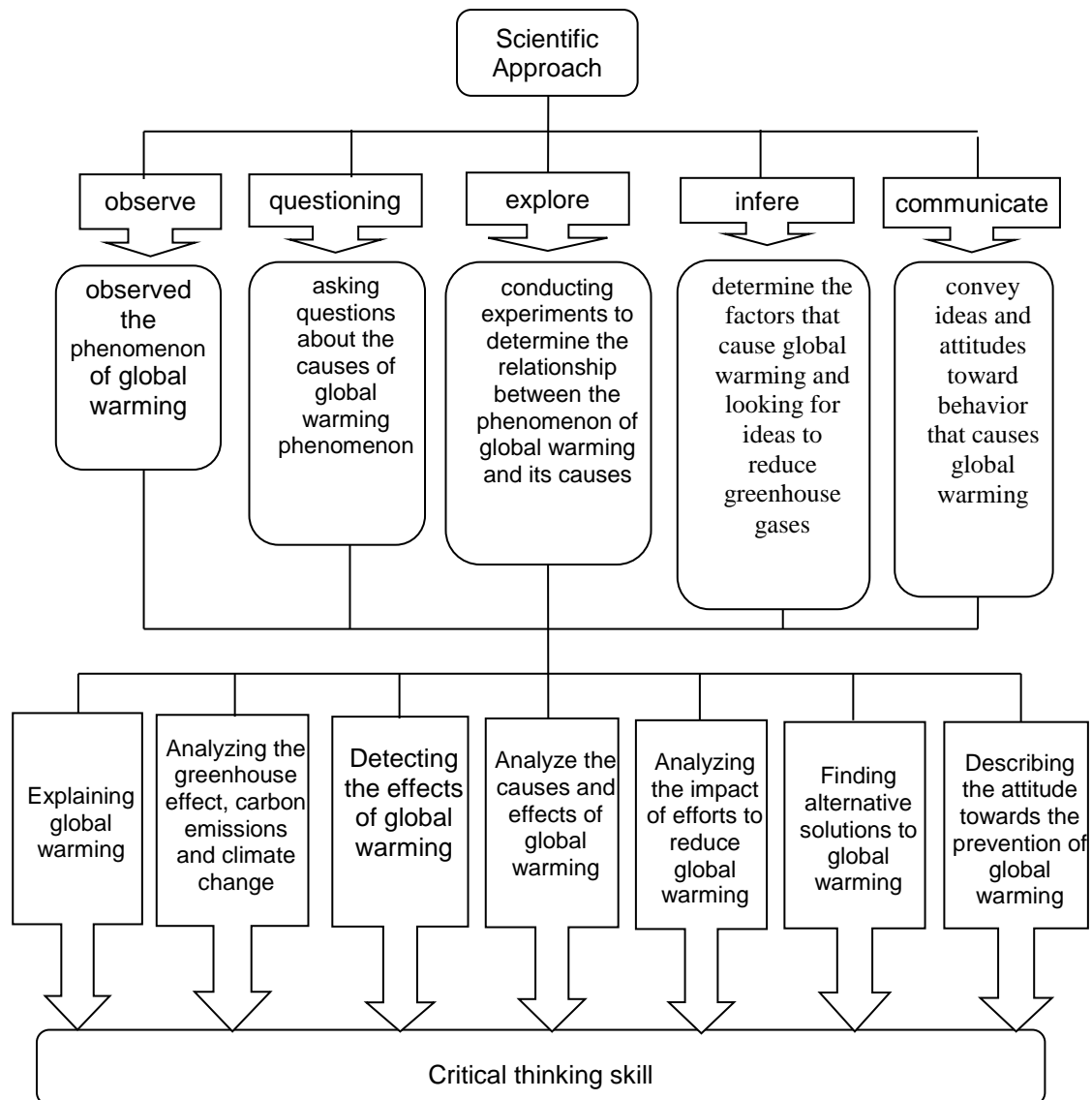


Figure 1. Assessment Models To Foster Critical Thinking Skills In Understanding Global Warming Looking For Ideas to Solving The Impact

At each stage of learning execute the assessment of global warming according to indicators of critical thinking using techniques and test forms adapted to the indicator. Indicators of critical thinking on global warming that are explaining global warming, analyzing the greenhouse effect, carbon emissions and climate change, detecting the effects of global warming, analyze the

causes and effects of global warming, analyzing the impact of efforts to reduce global warming, finding alternative solutions to global warming, describing the attitude towards the prevention of global warming.

IV. IV. CONCLUSSION AND RECOMENDATION

The results showed: (1) the students' understanding of the causes and effects of global warming for life is still low. There is a difference between knowledge of senior high school students who are studying global warming with no learning, but in junior high school students no difference. There is no difference in learning outcomes due to differences in gender and no interaction between the learning experience of global warming by gender, (2) teachers of physics require a model assessment of learning global warming using scientific approach to foster the idea of solving the problem of global warming, (3) critical thinking assessment on learning global warming using scientific approach, measure all indicators of critical thinking on every component of scientific approach using techniques and test forms adapted to the indicator.

It is necessary to find the relationship between students' knowledge of the impact of global warming with their attitude toward behaviors that contribute greenhouse gases.

It is need to develop models of global warming learning and media that can foster critical thinking skills in order to obtain innovative ideas to solving global warming impact.

ACKNOWLEDGMENT

Thanks submitted to the Directorate General of Higher Education, The Ministry of Research, Technology, and Higher Education who has been financed this research

REFERENCES

- [1] Wikipedia. Global Warming. Sumber: //id.wikipedia.org/wiki/pemanasan-global, accessed March 2011.
- [2] Skamp K. R, Boyes E, Stanisstreet M, Global warming responses at the primary secondary interface: 1 students' beliefs and willingness to act, Australian Journal of Environmental Education, 2009;25:15-30.
- [3] IPCC. *Climate change a threat to security, food and humankind* - IPCC report. (<http://www.theguardian.com/environment/2014/mar/31/climate-change-threat-food-security-humankind>) , accessed March April 2014.
- [4] Lee O, Lester BT, Ma L, Lambert J, Jean-Baptiste M. Conceptions of greenhouse effect and global warming among elementary students from diverse languages and cultures, Journal of Geoscience Education, 2007;55(2):117-25.
- [5] Marlon JR, Bartlein PJ, Carcaillet C, Gavin DG, Harrison SP, Higuera PE, Joos F, Power MJ, Prentice IC. Climate and human influences on global biomass burning over the past two millennia, Nature Geoscience 2008;1(10): 697-702.
- [6] Mitchell JFB, Johns TC. On modification of global warming by sulfate aerosols. Journal of Climate, 1997;10(2):245-67.
- [7] Taraneh Y, Sousan S, Mohammad RM, Seyed MS, Eddie B, Martin S, and Mirsaeed. A global warming: Knowledge and views of Iranian Students. Acta Medica Iranica, 2013; 51(3): 178-184
- [8] Hatcher, D. L., & Spencer, L. A. Reasoning and Writing: From Critical Thinking to Composition. 3rd. Ed. Boston: American Press, 2005
- [9] Robert D, Barbara L, Wendy W. Critical thinking framework for any discipline, International Journal of Teaching and Learning in Higher Education, 2006; 17(2): 160-166
- [10] Ennis, R. H. Critical Thinking. New Jersey: Prentice-Hall, 1996.
- [11] Mueller.J. The authentic assessment toolbox: Enhancing student learning through online faculty development. North Central College, 2005; 1(1): 1-7.
- [12] Chiu, S. & Chang, M. The Development of Authentic Assessments to Investigate Ninth Graders Scientific Literacy. International Journal of Science and Mathematics Education, 2005; 3: 117-140.
- [13] Wijayanti. Pengembangan autentik assesement berbasis proyek dengan pendekatan saintifik untuk meningkatkan keterampilan berpikir ilmiah mahasiswa, Jurnal Pendidikan IPA Indonesia, 2014; 3 (2), 102-108
- [14] Duron, R., Limbach, B., & Waugh, W. Critical Thinking Framework for Any Discipline, International Journal of Theaching and Learning in Higher Education, 2006; 17(2), 160-166.

THE NATIONALISM ATTITUDE ASSESSMENT OF STUDENTS OF STATE SENIOR HIGH SCHOOL 1 PAKEM SLEMAN

Aman

(Faculty of Social Science Yogyakarta State University)

Email: aman@uny.ac.id

ABSTRACT- The objective of this research is to find out nationalism attitudes of students of State Senior High School I Pakem Sleman, Yogyakarta Special Region. The research utilizes quantitative research method with evaluation research strategy. Senior High School students who become the sample of this research are 100 students of Grade XI who are considered as the representatives of students of grade X, XI, and XII. The research subjects are taken as proportional random sampling of all students of grade XI. The data collection utilizes questionnaire to assess students' nationalism attitudes. The data analysis technique utilizes quantitative descriptive data analysis technique with evaluation criteria. As a whole, the result of research indicates that nationalism attitudes of State Senior High School I Pakem Sleman which is based on the valuation of 100 students are 3.8 with good category. Male and female students' nationalism attitudes are 3.6 for males and 4.0 for females. Meanwhile, the valuation result of each aspect indicates that the aspects: proud of being Indonesian citizens 4.2, patriotism 4.1, sacrificing for the nation 3.6, proud of diverse cultures 3.8, respecting heroes' service 3.7 and prioritizing public interest 3.8. Thus, the result of this research indicates that students' nationalism attitudes of State Senior High School I Pakem Sleman are included in good category.

Keywords: *assessment, attitudes, and nationalism.*

I. INTRODUCTION

A social phenomenon that cannot be denied at the present time is the erosion of national and moral values among the younger generation. This phenomenon can be completely understandable given the complexity of variables that affect the integrity of the national existence of such negative effects of globalization and westernization penetrated into the joints of people who are still very fragile. At the time of shocks to the integrity of the nation was struck, components of the nation, including the younger generation is more preoccupied by the things that are practical and has been temporary. The noble values of the nation's culture bequeathed by the founding fathers have been uprooted. The younger generation will forget the identity and existence as a successor to the ideals of independence. In fact, indicators of independence needs to be built with strong national values, the spirit of nationalism, patriotism, religious prowess, collectivism, and the foundation of ultimate cultural values. However, the phenomenon says differently that our nation is dominated by schools individualism that ignores the principles of collectivism. Consequently, social issues and cultures are emerging everywhere. National identity "is traded" for the benefit of individuals or groups. No matter if the country loses or not, the important thing is that he/she is lucky and their group wins.

Regarding that, social science learning as the true means of national values and moral is often regarded as "additional" lesson. They are more obsessed with the hard sciences, and social sciences are considered less important. As a result, moral degradation among teenagers is great enough and undermines social values, which in essence is built. Loyalty to the identity and existence of the nation uprooted so that it is necessary to rethinking them in the form of building national values such as nationalism, national knowledge, and insight to the humanities and national awareness with a solid foundation of cultural values. Social learning requires a new paradigm in the context of character and national building, thus the national identity with noble values and dignity is rediscovered. A change in the mind set of learning social science is absolutely necessary because the importance of the socio-culture problems. It seems necessary

to build moral-scientific paradigm coherently-integrated, so that the meaningful learning in social science can take place in our education system.

At the time of the Indonesian faced a pile of problems caused by the various crises, the challenges in facing a globalization era characterized by openness and free competition is increasingly urgent. Inevitably, Indonesia must strive to improve the capabilities and competitiveness of human resources in the international arena. Within a relatively urgent situation, Indonesia should be able to prepare human resources that are professional, tough and ready to work. To implement these conditions, human resources of Indonesia need to have a stock of intellectual ability and the power of thought and the power of high innovation, also have knowledge, and the habit of applying good moral attitude. New ways of thinking and new breakthroughs should be introduced and created to overcome the problems of education in the present and in the future. To support it all, the quality of national education must be improved, so that the overall national educational goals can be achieved. Education reform with a variety of the segments is an imperative action (Zamroni, 2001: 158).

Hans Kohn, a historian who is very well known and most of his writings are on nationalism, provides terminology which is until today still relevantly used in the school learning, that is: *"nationalism is a state of mind in which the supreme loyalty of individual is felt to be due the nation state"*. That nationalism is an ideology which holds that the highest loyalty of the individuals must be submitted to the nation state (Hans Kohn, dalam A. Daliman, 2006). The concept of nationalism suggest that during centuries ago, the loyalty of individuals was not addressed to the nation state, but to the various forms of social power, political organizations, feudal kings, tribes, city state, the royal dynasty, a religious group or churches.

Similarly, the Indonesian republic founded together in the form of nation state according to the theories and principles of modern nationalism is very similar to that adopted by the United States. Construction of national unity was built on the concept of *Bhinneka Tunggal Ika* (pluralism) according to the pattern and its requirement is a product of history. Similarly, to build a strong determination of unity, our nation requires more than a quarter century by declaring three milestones, namely national revival with the founding of Boedi Oetomo in My 20th 1908, political manifesto by the Association of Indonesia (1925) and the declaration of proclamation of independence show that Indonesia has been released from the shackles of other countries (A. Daliman, 2006: 62).

Djoko Suryo (2005: 4) formulates rebuilding efforts of national attitudes those are: (1) nationhood attitude rooted in knowledge, understanding and experience of the concept of nation and nation state, (2) the concept of nation owned by the community today is basically a continuation of the concept of the nation according to the nationalism of the founders of the nation, (3) the vision of Indonesian nationalism during the national movement and the struggle for independence to be clearly formulated by the founder of the nation as the orientation of thought to realize the struggle for Indonesian independence from Dutch colonialism by establishing a unified state both the unity of the homeland as well as the language and culture, (4) characteristics and spirit of nationalism during the anti-colonial movements is the nature and spirit to build the unity of the people of the colonies from diversities into a united nation, where the motto *Bhinneka Tunggal Ika* from the Majapahit era was appointed as a slogan in an effort to realize the creation of building a nation that aspired, (5) the establishment of the Republic of Indonesia (the Declaration of Independence in 1945) as a manifestation of the success of nationalist movements in the period before independence and (6) independence revolution (1945-1950), became part of the embodiment of the ideals of nationalism maintains the nation state proclamation, and the journey of the country and the nation from the 1950s to the 2000s basically took place dynamically.

The spirit of nationalism in a nation-state nationalism inspired by five principles, namely: 1) unity, in a territory, nation, language, ideology, and doctrine of the state, political system or government, economic systems, security and defense systems, and culture policy; 2) freedom (liberty, freedom, independence), in religion, speech and oral and written, grouping and organizing; 3) equality, in the legal position, rights and obligations; 4) personality and identity, which has a self-esteem pride and compassion on the personality and identity of the nation that grew from and in accordance with the historical and cultural; 5) achievement is a dream for the welfare as well as the greatness and glorification of the nation (Sartono Kartodirdjo, 1999: 7-8).

Based on the Law of the Republic of Indonesia Number 20 of 2003 Article 3, explained that the national education serves to develop skills and character building and civilization of the

nation's dignity in the context of the intellectual life of the nation, is aimed at developing students' potentials in order to become a man of faith and fear of God Almighty, noble, healthy, knowledgeable, skilled, creative, independent, and become citizens of a democratic and responsible. Thus, national education has a very broad goal, not only related to academic prowess, but also other abilities such as religious, personal, and social. In the conception of history teaching objectives specifically realized it was more like a sense of history, nationalism, patriotism, insight humanities, in addition to academic skills. The problem of this study is as follows: how the nationalism attitudes of SMA N I Pakem Sleman.

II. RESEARCH METHOD

The method used is the evaluation of research or survey research evaluation strategy (Fernandes, 1984; Suharsimi Arikunto, 2004). The sample of the SMA students is about 100 students of Grade XI which are considered representing the students of Grade X, XI and XII. The subject of this study was collected through proportional random sampling from all of students of Grade XI. The data collection techniques used was questionnaire to assess nationalism attitude of the students. The indicators of nationalism attitude that are evaluated are: proud to be Indonesian, compassion for the homeland and the nation, willing to sacrifice for the sake of the nation, accept pluralism, proud of its diverse culture, appreciate the service of the heroes, and give priority to the public interest. Data analysis technique used is data descriptive analysis technique and using the evaluation criteria of Nana Sudjana (2005) with scale 1-5.

Table 1. Data Conversion Quantitative to Qualitative

Mean	Classification
> 4.2	Very Good
> 3.5 – 4.2	Good
>2.5 – 3.4	Fair
> 1.5 – 2.4	Poor
≤ 1.4	Very Poor

III. RESULT AND DISCUSSION

SMA Negeri 1 Pakem is a school under the auspices of the Ministry of Education and Culture. Since it was established in 1965, SMA Negeri 1 Pakem had undergone many changes, ranging from the name of the school to the existing infrastructure. As for the history and development of SMA Negeri 1 Pakem past until now is as follows: 1) in 1964 until 1965 named SMA III FIP IKIP Yogyakarta; 2) in 1966 until 1970 named SMA III IKIP Yogyakarta; 3) in 1971 until 1972 named SMA Percobaan III IKIP Yogyakarta; 4) in 1973 until 1974 named SM Pembangunan Yogyakarta; 5) in 1975 until 1986 named SMA Negeri III IKIP Yogyakarta; 6) in 1987 until 1999 named SMA Negeri Pakem Yogyakarta; 7) in 1996 until 2003 named SMU Negeri 1 Pakem Yogya, and started from 2003 until now named SMA Negeri 1 Pakem (Profil SMA N 1 Pakem 2015). SMA Negeri 1 Pakem commemorate the anniversary every year on August 13. The school is ideally located, as it is easy to reach by students and is located close to the highway. This is the physical potential to support the learning process. Location of SMA Negeri 1 Pakem precisely on Jl. Kaliurang Km. 17,5 Pakem, Sleman, Yogyakarta.

Profile of SMA Negeri 1 Pakem has the vision, mission, and goals that are clear and measurable. The vision of SMA Negeri 1 Pakem is superior in achievement, noble to have a global perspective that is based on the nation's cultural roots. Missions of SMA Negeri 1 Pakem are: 1) creating a learning culture for all member schools, 2) create a culture of achievement and pride for the entire school community, 3) creating efficiency and effectiveness of teaching and learning process, 4) improve achievement and school rankings at national level, 5) building disciplines of the students, dedicated, honest, and uphold human dignity, 6) building human that orderly practice her/his faith to always devoted to God Almighty, 7) create a culture of competing on a global level for the entire school community, 8) developing human that is creative, innovative, and competitive for the whole school community in a dignified and respectful in manner, 9) strengthen invention of the nation's cultural values to shape the character of the nation as well as building and respect for local knowledge, and 10) creating a harmonious atmosphere of school life, harmonious, and balanced (Profil SMA N I Pakem Sleman 2015).

The purpose of SMA Negeri 1 Pakem are: 1) prepare learners devoted to God Almighty and noble ; 2) create the learners to achieve high academic achievement; 3) prepares students to become human that have personality, smart, qualified, and good in the field of sports and arts; 4)

to equip students to have the skills of information technology and communication and is able to develop themselves independently; 5) instill learners tenacious attitude and persistent in compete, adapt to the environment, and develop an attitude sportively; and 6) provide students with science and technology in order to compete and continue to pursue higher education.

SMA Negeri 1 Pakem is one of the flagship high school whose existence was already quite old and proved to be able to contribute to the intellectual life of the nation, it is located on Jl. Kaliurang Km. 17,5, Pakem, Sleman, Yogyakarta. The condition is strategic and conducive school as a place of learning. It can be seen from the location that it is located near the highway so it is easy to reach by public transport. In addition, the atmosphere was not too crowded so as to enable the implementation of teaching and learning to run smoothly and quietly. SMA Negeri 1 Pakem already equipped with some supporting infrastructure of teaching and learning process, including school building consisting of classrooms or study hall, office space, support and field space which is used for ceremonial activities, sports and extracurricular for implementation.

SMA Negeri 1 Pakem has teachers about 36 people, who are mostly qualified S1 (Bachelor) and few teachers hold Master Degree (S2). Most teachers are civil servants, and some teachers are still non-civil servants. Each of these teachers teaches based on their expertise. In addition, there are also some teachers to guide the students. While the employees at SMA Negeri 1 Pakem are nine members, five people of Administrative Affairs, one person of librarian, 2 people of public servants (janitor, parking, school kitchens) and one person of night guard.

Based on the results of this evaluation, the results of research regarding the attitude of nationalism is as follow. The overall student's nationalism attitude of SMA N I Pakem Sleman according to the assessment of 100 students is 3.8 which fall in Good category. There are differences of the students' nationalism between the male students and female students, the male one is 3.6 and the female one is 4.0. Whereas the assessment results of nationalism attitude in each aspect shows that aspects of: proud to be Indonesians is 4.2, love of the homeland and the nation is 4.1, willing to sacrifice for the nations is 3.6, accept pluralism is 3.6, proud of the diverse cultures is 3.8, appreciate the services of the heroes is 3.7, and prioritizing public interest is 3.8. Therefore, educators must develop strategic programs to provide students with a cross-cultural skills including embedded code nationalism. The education should be developed to accommodate the values of the local community (Hannerz, 2009). Thus, effective school management must involve all components in school together to achieve the school's vision in leading a school that has a sense of history and nationalism students (Nunuk Suryani, 2013). The results are ratings nationalism SMA N I Pakem Sleman is as follows.

Table 2. The Result of Students' Nationalism Attitude

No.	Assessment Aspects	Mean	Classification				
			VP	P	F	G	VG
1	Proud to be Indonesians	4.20					√
2	Love of the homeland and the nation	4.10				√	
3	Willing to sacrifice for the nations	3.60				√	
4	Accept pluralism	3.60				√	
5	Proud of the diverse cultures ragam	3.80				√	
6	Appreciate the services of the heroes	3.70				√	
7	Prioritizing public interest	3.80				√	
Total		26.6					
Mean		3.80					

N=100

VP=Very Poor; P=Poor; F=Fair; G=Good; VG=Very Good

Based on the results of students assessment of the implementation of components and outcome indicators of teaching history in SMAN 1 Pakem, Sleman indicates that the attitude of the students included in the classification of nationalism either. Nationalism had a mean score of 3.80. This indicates that students are already good nationalism. There is one indicator that received a score of 4.20 with very good categories, the name of the indicator is proud to be Indonesian, while six other indicators ranged from 3.60-4.10 with good category.

Table 3. The Result of Nationalism Attitude Assessment of Female Students

No.	Assessmen Aspects	Mean	Classification				
			VP	P	F	G	VG
1	Proud to be Indonesians	4.38					√
2	Love of the homeland and the nation	4.20					√
3	Willing to sacrifice for the nations	3.72				√	
4	Accept pluralism	3.74				√	
5	Proud of the diverse cultures ragam	4.16				√	
6	Appreciate the services of the heroes	3.70				√	
7	Prioritizing public interest	4.18				√	
Total		28.08					
Mean		4.00					

N=50

VP=Very Poor; P=Poor; F=Fair; G=Good; VG=Very Good

Based on the results of student assessment of the implementation of components and indicators of learning outcomes in the history of SMA N I Pakem Sleman, it shows that nationalism of female students included in good classification. Nationalism had a mean score of 4.00. This indicates that the nationalism of female students has been good. There are two indicators that obtained a score of 4.38 and 4.20 with very good categories, they are proud to be Indonesian and love of the homeland and the nation, while the other five indicators ranged 3.70-4.18 with good category.

Table 4. The Result of Nationalism Attitude Assessment of Male Students

No.	Assessment Aspects	Mean	Classification				
			VP	P	F	G	VG
1	Proud to be Indonesians	4.02				√	
2	Love of the homeland and the nation	4.00				√	
3	Willing to sacrifice for the nations	3.48				√	
4	Accept pluralism	3.46				√	
5	Proud of the diverse cultures ragam	3.44				√	
6	Appreciate the services of the heroes	3.70				√	
7	Prioritizing public interest	3.42				√	
Total		25.2					
Mean		3.60					

N=50

VP=Very Poor; P=Poor; F=Fair; G=Good; VG=Very Good

Based on the results of student assessment of the implementation of components and outcome indicators of teaching history in SMA N 1 Pakem, Sleman indicates that the nationalism attitude of male students fall in good classification. Nationalism had a mean score of 3.60. This indicates that the nationalism attitude of male students had been good. All indicators are in a good category which ranging from 3.70-4.02.

IV. CONCLUSION

Based on the results and discussion in this study, it can be concluded as follows. The overall student's nationalism attitude of SMA N I Pakem Sleman according to the assessment of 100 students is 3.8 which fall in good category. There are differences of the students' nationalism between the male students and female students, the male one is 3.6 and the female one is 4.0. Whereas the assessment results of nationalism attitude in each aspect shows that aspects of: proud to be Indonesians is 4.2, love of the homeland and the nation is 4.1, willing to sacrifice for the nations is 3.6, accept pluralism is 3.6, proud of the diverse cultures is 3.8, appreciate the services of the heroes is 3.7, and prioritizing public interest is 3.8. Therefore, the results of this study indicate that nationalism attitude of SMA N 1 Pakem, Sleman students included in good category. This result shows that nationalism attitude of female students is higher than the nationalism attitude of male students. Thus it can be justified that the nationalism attitudes of SMA N 1 Pakem, Sleman included in a good category. This good result is the basis for school

improvement related to both academic and school quality and other sectors, especially with regard to the teaching of history that emphasizes the efforts of investing national values.

V. SUGGESTIONS

Based on the evaluation above, the suggestions are formulated as follows.

1. Giving suggestions to the school supervisors to utilize the result of this study in assessing managerial ability of the headmasters in improving the quality of school functions both in academic sector as well as in non-academic sector, especially related to substantive character education.
2. Giving suggestions to the History teachers to continue improving their ability especially in teaching History so that the goal of the learning which substantively related to attitude can be well-achieved.
3. To the other researchers, the result of this study focuses on nationalism area, and it still can be develop with a broader research area, for example regarding to the attitude of defending the country, knowledge of history, a sense of history and other relevant attitudes to the purpose of history learning.

REFERENCES

- [1] Anastasius Daliman.(2006). "Harmonisasi antara nasionalisme dalam kehidupan bernegara dan beragama", in *kearifan sang profesor*. Yogyakarta: Universitas Negeri Yogyakarta Press.
- [2] Djoko Suryo.(2005). "Paradigma sejarah di Indonesia dan kurikulum sejarah", dalam *makalahseminar nasional dan temu alumni program studi pendidikan sejarah Program Pascasarjana Universitas Sebelas Maret Surakarta*.Surakarta: PPS UNS.
- [3] Fernandes, H.J.X. (1984). *Evaluation of educational programs*. Jakarta: National Educational Planning Evaluation and Curriculum Development.
- [4] Hannerz, U. 2009. Cosmopolitans and Local in World Culture.*Journal Theory , Culture, and Society* 7 (2)237-251.
- [5] Nana Sudjana. (2005).*Penilaian Pendidikan*. Bandung: Rosda Karya.
- [6] SMA N I Pakem.(2015). *Profil SMA Negeri I Pakem Sleman Tahun 2015*. Sleman: SMA N I Pakem Sleman.
- [7] Sartono Kartodirdjo. (1999). *Multidimensi pembangunan bangsa etos nasionalisme negara kesatuan*. Yogyakarta: Kanisius.
- [8] Suharsimi Arikunto. & Cepi Safruddin AJ. (2004). *Evaluasi program pendidikan, panduan teoritis praktis bagi praktisi pendidikan..* Jakarta: Bumi Aksara.
- [9] Suryani, N. 2013. Manajemen Pembelajaran Berbasis Kultur di Sekolah. *Jurnal Paramita* 23 No.2 2008-219.
- [10] Zamroni. (2001). *Paradigma pendidikan masa depan*. Yogyakarta: Bayu Indra Grafika.

THE DESIGN OF FORMATIVE ASSESSMENT BY INQUIRY BASED LEARNING IN IMPROVING STUDENTS' SELF-REGULATION

Asih Sulistia Ningrum¹, Chandra Ertikanto²

¹ Teacher of Madrasah Al Hidayah Raman Utara, Lampung

² Master of Magister of Physics Education, Lampung University
acihtya@gmail.com¹

Abstract—The purposes of this research are: (1) To determine the need of teacher and students' formative assessment that is suitable to the increase of self-regulation skills by inquiry based learning; (2) To analyze the learning process of physics (3) designing a model of formative assessment inquiry based in improving self-regulation students'. The research was conducted in three high school in East Lampung. This research to use descriptive method by cluster sampling. The procedure of research that has been done include (1) the initial research and data collection, (2) planning, and (3) developing the initial product. The result of the research show that: (1) teachers and students' high school need assessment model such as of formative assessment inquiry based in improving students' self-regulation in learning physics through the concepts, principle and laws of physics; (2) The implementation of inquiry learning has not been implemented optimally and required model of formative assessment inquiry based corresponding to the learning need of students' as in improving students' self-regulation in the understanding of physics concept better; (3) Design of formative assessment at developing based on the preliminary study. Early stage a model of formative assessment inquiry based consist of several step are: orientation, formulating the problem, hypothesis, analyzing data, and conclusion.

Keywords: *formative assessment, self-regulation, inquiry*

I. INTRODUCTION

The achievement of the learning process of students', teachers need to do the assessment activity. Assessment of conducted during the learning process by providing feedback (feedback) and facilitate student self-assessment to monitor the progress and improve the teaching and learning process. Feedback is done at end of the learning as a value and description in the report, does not provide an opportunity for students' improve their learning. Therefore, the feedback is not only done at the end of learning, but also during the learning process. One assessment that provide feedback once the skill to assess the self is a formative assessment.

A formative assessment is formally are designed or planned activities before the teaching, consists of three phases: eliciting, interpreting and acting on information assessment [1]. Formative assessment have the concrete step, such as (a) to anticipate and led to the idea of students', (b) evaluate the idea of students', and (c) preparing the next step in the instructions that explain the idea of students" and support student learning [2]. At formative assessment process, where teacher can manage the attention and think response students' to assisting students' in improving learning [3]. Formative assessment is not a test or a tool but a process with the potential to support learning outside the school with developing individual learning strategis [4]. In addition, formative assessment in the classroom is a process of instruction that have the strong potential for the task can provide information that supports learning assessment into the design of learning process [5]. The investigation formative assessment in the form of question to use determine a students' knowledge before the learning and determining the methods to teach the learning objectives [6]. The purpose offers feedback is to help students' achieve success desired performance [7]. The use of feedback is closely related to student engagement and peer assessment [8]. Formative assessment through feedback impact on student learning activity [9].

A self-regulation is an activity of self-control leads to mastery of specific techniques to actual learning, such as self instruction, compare, focus, strategy tasks, connect and seek helping [10]. Self-regulation (SR) is a process of self-assessment in which students' tasks represent, plan how to implement, monitor and assessing whether the implementation is sufficient, able to overcome difficulties and emotional that normally arise, assessing performance and make the relationship conclusion about the result [11]. Self-regulation is defined as a form of learning individually as includes independence that is starts from social learning, such as seeking help from peers, coaches, and teachers [12]. Construction of self-regulation will be specifically related to the emotional experience and expression [13]. Thus, students' with higher levels of self-regulation show a higher level anyway to get involve in demonstration [14].

Formative assessment by components of related to the ability of self-regulation is the feedback, self-assessment and peer assessment. This assessment has not been fully to use by the teacher in learning. Self and peer the assessment are the two most common forms used in formative assessment [15]. If the assessment is carried out with the involvement of inquiry activity can help teachers assessing students' affective aspects. Also active by students' through are self and peer assessment can improving students' engagement in facilitating individual feedback and make students' more responsible about their learning progress. A physics is one of the subject that are considered difficult by most students', learning physics are faced with a variety of concept and formula. Students' who are to use independently will not have trouble learning. Positive emotion will show a positive assessment of the value of the task and/or the results of achievement, and positive emotion should improve cognitive and motivational processes inherent in self-regulation [16].

One of the learning activities used to measure students' activity and to the controlling of ability by students' as independently is inquiry learning activity. The concept of inquiry learning is a pedagogical approach to the character of the students' in learning activity in driving ability through scientific knowledge that the content and process skill [17]. An inquiry group can create a collaboration based on the concept of knowledge [18]. Teachers provide opportunities to more accurately gauge students' knowledge through scientific idea and the process involved in the class [19]. Inquiry based instruction requires a new way of student engagement in the learning process and that is why teacher need to look at important factors as changes [20]. Thus, inquiry based learning is very good as a center learning activity [21]. However, the idea of such an inquiry is still rarely developed back through reflection science [22].

Based on the explanation above, the purpose of this study are: (1) To determine the need of teacher and students' formative assessment that is suitable to the increase of self-regulation skills by inquiry based learning; (2) To analyze the learning process of physics (3) designing a model of formative assessment inquiry based in improving self-regulation students'.

II. RESEARCH METHOD

This research method to use the method of research and development. In this developing research generated design of formative assessment by inquiry based to improving students' self-regulation in the learning high school physics. The development procedure refer to a procedure conducted research [23], which include the step of: (1) research and data collection beginning; (2) planning; and (3) developing the initial product.

Method of data collection and analysis of need carried out to use sampling techniques such as cluster sampling, that the sampling carried out on the sampling unit, where the sampling unit consist of one group (cluster). This research was conducting in three high school in East Lampung. The research subject class X SMA numbered 40 students'. Research subjects three teacher of physics. That matter aim to obtain a description of the analysis of the need of the developing of formative assessment by inquiry based to improving students' self-regulation in the learning high school physics. Meanwhile, for the design developing of formative assessment by inquiry based to improving students' self-regulation is done through the study of literature. The new design developed until the early stage of developing a product to produce the prototype I by model formative assessment through cycle of activity inquiry based learning.

The data collection phase initial conducted by questionnaire analysis of the need of teacher and students'. Questionnaire data analysis technique performed by several stage; (1) giving the code of the distribution of data to the questionnaire; (2) The data tabulation is based on the classification made; (3) the analysis of qualitative data that outline as well as connecting the data

with information related to the research of focus; and (4) interpret of the result at thorough analysis and making conclusion.

III. RESULT AND DISCUSSION

The development of new research conducted up to the third stage of the research procedure [24] is planning to developing of the initial product, so that the results of the design is successfully developing a prototype I called. However, before producing the first prototype, the researcher conducted the analysis of the need of teacher and students'.

The results of questionnaire analysis disclosure need of teacher, gained P1 to P3 answer: as much as 66.67% of the teacher are already to use the assessment of instrument learner outcome learning end of each chapter analyzing the answer P5 total of 100% of teacher agree that organized inquiry learning activity when conducting by the assessment.

Analysis of the questionnaire answer reveal P6 to P18 stage of inquiry learning activity are: the orientation stage, 33.33% of the teacher do not direct to lead a study plan to motivate of interest students' learning. The second stage of the formulating problem P7 to P9, as much as 66.67% of teachers leave are question related to current example of learning physics of this case students' have not been able to formulate the problem individually. The third phase of the hypothesized P10: as much as 33.33% of teacher do not lead students' to this proces temporary answer. The fourth stage of the analyzing data P11 to P17, there are 33.33% of the teachers encourage students' to comparing at answer according to the theory. The fifth stage concludes P18, 66.67% of teacher give students' with an opportunity to conclude of the data which have been obtained.

The analysis an answer P20, P21, and P22, as many as 66.67% of the teacher are still experiencing trouble making a instrument of the learning outcome. Much as 100% of teacher in need of assessment sheet for measuring result of learning students' and the need to developing of formative assessment by inquiry based to improving students' self-regulation. Based on the analysis of an answer is known that most teacher have not been up applying inquiry learning activity because for teacher are still applying the conventional method. Implementation by of activity inquiry learning to be able controlling the students' self-regulation. Self-regulation is very important for the students' to carry out academic and a key factor in understanding, intervening, and experiencing academic difficulty [25]. So the researcher concluding that teacher need model of formative assessment by inquiry based to improving students' self-regulation.

Whereas, the results of questionnaire the disclosure need of high school students' can be known that a student need to once the assessment of the learning outcome. The average analysis of the answer P1 to P3, many as 62% of students' need a formative assessment for improving of physics concept is a carefully. Into the bargain, a students' of need are physics learning which is interesting and fun by mean of inquiry learning activity. The average analysis of the answer to P5 to P18, 60% of students' need at inquiry learning activity in problem solving, orientation, formulate hypotheses, analyzing data and concluding when the learning activity increase of the ability, controlling, and monitoring the knowledge of at each individual.

The analysis of an answer P19, P20: as much as 83% the students' need a result value after a conducting an investigation and 100% the students' need a sheet of the assessment for measuring learning outcome by the obtained. Based on the analysis of this answer, it is known there are students' which is have difficulty in understanding the concept, law and principles of physics through by the inquiry learning activity for establish and improving self-regulation of this each individual. A inquiry based instruction requires a new way of student engagement in the learning process and that is why teacher need to look at important factor as change [26]. So the researcher concluding that it need to be developing of formative assessment by inquiry based in improving students' self-regulation.

A formative assessment model that will be developing more are orientation to the students' self-regulation approach through is inquiry, so that the students' will participate in the cognitive by are learning environment, it is based on the result of a development analysis. Once the researchers obtain of descriptive conclusions about the need for development, researcher follow in by step procedure are developing early product. The stage that have been conducted by researcher to produce model of formative assessment by inquiry based to improving students' self-regulation explaining as follow.

The first phase, which is constructing the basis of reasoning or framework. Researcher create a framework to explain the process of learning and assessment through the model to be developed. The second stage, which bring about a study literature to determine the formative assessment process by that teacher, opinions of teachers and students' on model of formative assessment by inquiry based, and is support by a study theory of contributing to the concept from model will be developing.

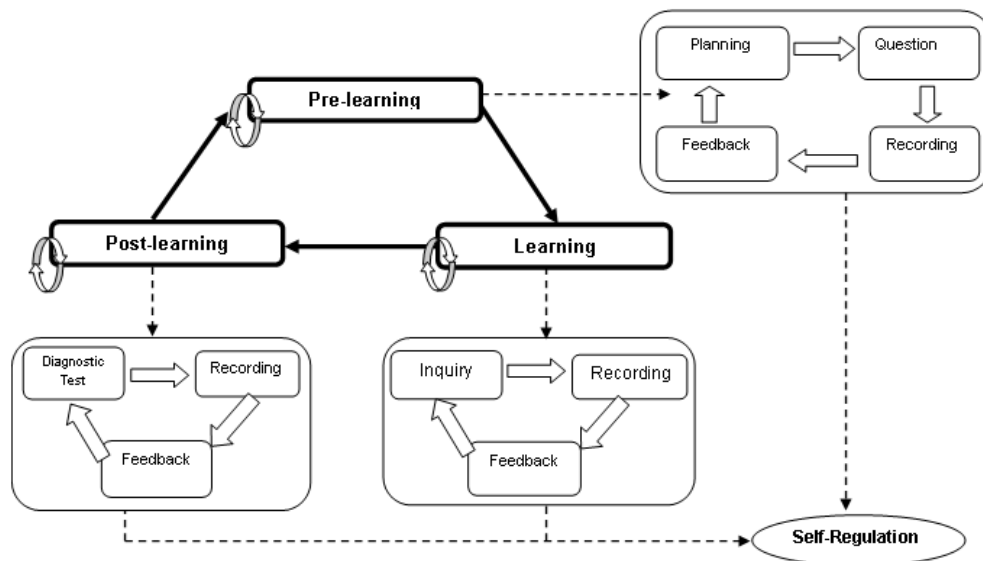


Figure 1. The Design Of Formative Assessment By Inquiry Based

Furthermore, the development of planning stage to model of formative assessment by inquiry based were developing to assist teacher in implementing at formative assessment so as to improving students' self-regulation in at learning. The model successfully developed the researcher focused on formative assessment by inquiry based. Such an approach would emphasize on this students' self-regulation at learning activity. Figure 1, researcher presented the formative assessment cycle through inquiry based at learning activity.

The model of formative assessment consist from a prelearning, learning and post-learning. The learning activity in the by inquiry activity conducted by researcher at through several stage, including orientation the problem, formulate hypothesis, analyze data, and concluding. Planning researcher in stage pre-learning and post-learning at contain a quiz with question adapted to all component that support by continuity of the learning device. In providing feedback at each stage to know the extent of the feedback that will be developing in accordance with aspect of the formative assessment. The result of formative assessment design are designed the researcher through inquiry approach to in providing feedback every learning process, students' can improving self-regulation in particular the ability of understanding the cognitive independently.

IV. CONCLUSION

Based on the research this purpose and the above presentation, it can be concluding that: (1) teacher and students' high school need assessment model such as of formative assessment inquiry based in improving students' self-regulation in learning physics through the concept, principle and laws of physics; (2) The implementation of inquiry learning has not been implemented optimally and required model of formative assessment inquiry based corresponding to the learning need of students' as in improving students' self-regulation in the understanding of physics concepts better; (3) Design of formative assessment at developing based on the preliminary study. Early stage a model of formative assessment inquiry based consist of several step are: orientation, formulating the problem, hypothesis, analyzing data, and conclusion.

ACKNOWLEDGMENT

This project would not be possible without the help from members of our team who are not author on this paper Erlida Amnie, Luthfi Riadina, Saiful Ali Nurdin, Novinta Nurul Sari, Nurul Ulil

Amri. Arti Mela Agustin, tuti Widyawati, Imas Estiana Galih, Ferico Chandra, M. Iwan, Farid Bayu Nirwana, and Yuda Seta Mahendra.

REFERENCES

- [1] I. Ali, and H. M. Iqbal, "Effect of Formative assessment on Students' Achievement in Science," *Word Applied Sciences Journal. Pakistan*, Vol. 26, pp. 677-687, November 2013.
- [2] J. L. Sabel, C. T. Forbes, and L. Zangori, "Promoting Prospective Elementary Teachers' Learning to Use Formative Assessment for Life Science Instruction," *The Association for Science Teacher Education. USA*, vol. 26, pp. 419-445, May 2015.
- [3] E. M. Furtak, "Linking a Learning Progression for Natural Selection to Teachers' Enactment of Formative Assessment," *Journal Of Research in Science Teaching. Colorado*, vol. 46, pp. 1181-1210. October 2012.
- [4] I. Clark, "Promoting Prospective Elementary Teachers' Learning to Use Formative Assessment for Life Science Instruction," *Educational Psychology Review. USA*, January 2016.
- [5] I. Clark, "Formative Assessment: Translating High-Level Curriculum Principles Into Classroom Practice," *Department of Communications. Nagoya University of Commerce and Business (NUCB). Japan*, vol. 26. Pp. 91-114, 2015.
- [6] N. Bulunuz, M. Bulunuz, F. Karagos, and O. F. Tavsanh, "Achievement Levels of Middle School Students' in the Standardized Science and Technology Exam an Formative Assessment Probes: A Comparative Study," *Journal of Education in Science. Environment and Health (JESEH), Uluvol. 2*, pp. 33-50, 2016.
- [7] S. Tosati, N. Lawthong, and S. Suwanmonkha, "Development of An Appreciative Inquiry and Assessment Processes For Students' Self-Knowing And Self-Development," *Procedia Social and Behavioral Sciences. Elsevier. Thailand*, vol. 191, pp. 753-758, 2015.
- [8] A. havnes, K. Smith, O. Dysthe, and K. Ludvigsen, "Formative Assessment and Feedback: Making Learning Visible," *Studies in Educational Evaluation. Norway*, Vol. xxx, pp. xxx-xxx, 2012.
- [9] Nahadi, H. Firman, and J. Farina, "Effect of Feedback in Formative Assessment in the Student Learning Activities on Chemical Course to the Formation of Habits of Mind," *Indonesian Journal of Science Education. Indonesia*, Vol. 4, pp. 36-42, April 2015.
- [10] B. J. Zimmerman, "From Cognitive Modeling to Self-Regulation: A Social Cognitive Career Path," *Educational Psychologist. New York*, Vol. 46, pp. 135-147, 2013.
- [11] E. Panadero, J. A. Tapia, and J. A. Huertas, "Rubrics and Self-Assessment Scripts Effects on Self-Regulation, Learning and Self-Efficacy in Secondary Education," *Learning and Individual Differences. Elsevier. Spain*, Vol. 22, pp. 806-813, April 2012.
- [12] B. J. Zimmerman, and D. H. Schunk, *Handbook Of Self-Regulation Of Learning And Performance*. Routledge Taylor & Francis Group: New York and London, 2011.
- [13] D. J. Bridgett, K. B. Oddi, L. M. Laake, K. W. Murdock, and M. N. Bachman, "Integrating and differentiating aspects of self-regulation: Effortful control, executive functioning, and links to negative affectivity," *Emotion. Department of Psychology. Northern Illinois University*, Vol. 13, pp. 47-63, 2013.
- [14] J. Chin-Yuan Sun, and R. Rueda, "Situational Interest, Computer Self-Efficacy and Self-Regulation: Their Impact on Student Engagement in Distance Education," *British Journal of Educational Technology. California*, Vol. 43, pp. 191-204, 2012.
- [15] M. Aydeniz, and M. A. Gilchrist, "To use Self-Assessment to Improve College Students' Engagement and Performance in Introductory Genetics," *Necatibey Faculty of Education Electronic Journal of Science and Mathematics. USA*, Vol. 7, pp. 1-17, October 2013.
- [16] Villavicencio, T. Felicidad, and A. B. I. Bernardo, "Positive Academic Emotions Moderate the Relationship Between Self-Regulation and Academic Achievement," *British Journal of Educational Psychology*, Vol. 83, pp. 329-340, 2013.
- [17] T. L. Bunterm, K. K. Lee, N. L. Srikoorn, P. Vangpoomyai, J. Rattnavongsa, and G. Rachahoon, "Do Different Levels of Inquiry Lead to Different Learning Outcomes? A Comparison Between Guided and Structured Inquiry," *Internasional urnal of Science Education*, Vol. 36, pp. 1937-1959, 2015.
- [18] D. Dumitru, "Communities of Inquiry, A Method to Teach," *Procedia Social and Behavioral Sciences. Elsevier. Romania*, Vol. 33, pp. 238-242, 2012.
- [19] M. Aydeniz, D. F. Cihal, S. C. Graham, and L. Retinger, "To use Inquiry based Instruction For Teaching Science to Students' With Learning Disabilities," *International Journal of Special Education. Tennessee*, Vol. 27, No. 2, 2012.
- [20] J. Dostal, and M. Klement, "Inquiry based Instruction and Realiting Appeals of Pedagogical Theories and Pratices," *Procedia Social and Behavioral Science. Elsevier. Czech Republic*, Vol. 171, pp. 648-653, 2015.
- [21] A. Veloo, S. Perumal, and R. Vikneswary, "Inquiry based Instruction, Students' Attitudes and Teachers' Support Towards Science Achievement in Rural Primary Schools," *Procedia Social and Behavioral. Elsevier. Malaysia*, Vol. 93, pp. 65-69, 2013.
- [22] S. J. Zuiker, and J. R. Whitaker, "Refining Inquiry with Multi-Form Assessment: Formative and Summative Assessment Functions for Flexibel Inquiry," *International Journal of Science Education*, Vol. 36, pp. 1037-1059, 2014.
- [23] M. D. Gall, J. P. Gall, and W. R. Borg, *Educational research: An introduction (7th ed.)*, Boston, MA: Pearson Education, 2003.
- [24] M. D. Gall, J. P. Gall, and W. R. Borg, *Educational research: An introduction (7th ed.)*, Boston, MA: Pearson Education, 2003.

- [25] M. M. Mecklelland, and C. E. Cameron, "Self-Regulation in Early Childhood: Improving Conceptual Clarity and Developing Ecology Valid Measures," Vol. 6, pp. 136-142, 2012.
- [26] J. Dostal, and M. Klement, "Inquiry based Instruction and Realizing Appeals of Pedagogical Theories and Practices," Procedia Social and Behavioral Science. Elsevier, Vol. 171, pp. 648-653, 2015.

Exploring the Use of One Meeting Theme-Based Extended Response

A Practical Critical Thinking Assessment Tool for Classroom Practices

Ayu Alif Nur Maharani Akbar¹, Rahmad Adi Wijaya²

¹ Undergraduate student of English Language Teaching (Faculty of Letter, State University of Malang)

² Undergraduate student of English Language Teaching (Faculty of Letter, State University of Malang)

Abstract— Despite the recently growing interest of implementing critical thinking assessment in classroom, some teachers still contend that assessment of this kind is burdensome for them since they acknowledge it as intricately convoluted process. Indeed, it is also disputed whether their students will be able to carry out the tasks. In addition, critical thinking assessment is considered as impractical and unfeasible due to the limited time allotments given at school. This paper is aimed at discussing how actually a model of theme-based extended response can debunk the imprudent belief about the notions of higher order thinking assessment. Designing a one meeting worksheet consisting of theme-based extended response instructions developed based on scientific approach in Curriculum 2013 is not that complicated for teachers. Moreover, this one meeting theme-based response is a practical authentic assessment for classroom practices, easily attainable by teachers, that is intended to augment their students' higher order thinking skills.

Keywords: *critical thinking assessment, extended-response, higher order thinking skills, theme-based*

I. INTRODUCTION

The rapid growth of modern websites, online literature, and other digital technologies should make us more aware that we may not unquestioningly swallow all information we access. Media that we are exposed to, most likely, is prone to provide us with information might be influenced by subjective perspectives, biases and even bogus claims. Hence, one's ability to think critically and make a sound judgment about certain issues is indispensable. Ennis(1993) states that critical thinking has become the goal of education within this century. The critical thinking ability is crucial for someone to make reasonable and defensible decision toward an issue, evaluate different points of view, analyze existing believes and paradigms, and implement their knowledge into practices (Symth, 2000).

II. THE NOTION OF CRITICAL THINKING

Pertaining to our ability to think, critical thinking pervades our thought process with shrewd and sound judgment to weigh up the merits of something by employing logically acceptable standards. Yet, it is worth noting that being critical is not simply about being fault-finding. There are many skills one can develop to be critical instead of getting fixated on merely an evaluative skill. Critical and creative thinking are both inseparable as they two give ample space for creativity and imagination. Thus, when it deals with creativity and imagination, one cannot argue that critical thinking is not applicable in classroom context due to its complexities, some teachers may think. It hinges upon how well the teachers comprehend the importance of critical thinking for their students' learning and how critical and creative the teachers are to figure out ways to exert it in their teaching; students need their teachers to drive them into their utmost development of critical thinking.

Since early 1980s, the attention to design more critical thinking instructions has increased significantly (Ennis, 1993). This interest leads the development of assessment encouraging

students' higher order thinking skills-critical thinking assessment. Referring to bloom taxonomy, higher order thinking skills consist of several stages of thinking: analyzing, synthesizing, evaluating, and creating. However, the definition of critical thinking itself is still vague. Ennis (1993) provides clearer definition of critical thinking skills; he states that critical thinking is reasonable reflective focus on deciding what to believe or what to do. This definition highlights the goals of critical thinking: being reasonable and reflective. Conceptually, critical thinking can be further expounded by referring to one's skills to discern false inferences and logical fallacies, distinguish bias from fact and so forth. Indeed, critical thinking compels our thoughts to be purposeful and self-regulating (Scriven & Paul, 1987). Having clear aims and intellectual self-assessment in mind can drive problem-solving and decision-making. In a nutshell, critical thinking can characterize one's competencies in purposefully employing systematic reasoning and logic to make good decisions in order to solve problems.

III. ELEMENTS AND STAGES OF CRITICAL THINKING

Names like Benjamin Bloom and Lorin Anderson may have reached our ears often through their work on a classification of intellectual behavior levels. Bloom's taxonomy on cognitive skills has formed the bedrock of teaching and learning, and also assessment principles. In addition, high-order thinking skills, according to the taxonomy, involve several layers of cognitive domains: analyzing, synthesizing, evaluating and also creating.

IV. AUTHENTIC ASSESSMENT AND CRITICAL THINKING

When it comes to assessment, most teachers are highly aware that it becomes an inseparable part of their teaching activities. However, it is likely that the form of assessment employed for both standardized-testing and classroom assessment for as long as most teachers can recall is the multiple-choice test format. Realizing its drawbacks in terms of the constructs it seeks to measure though in many ways it also provides great advantages like its nature that is highly quantifiable, many test-designers and educators have raised deep concerns about seeking assessments that can resemble instructional activities in classroom and can reveal students' competences in the truest sense. Stiggins (1991) states that alternative assessments entail any method of finding out what a student knows or can do and are typically authentic since it is based on activities that represent classroom and real-life settings.

Indeed, assessment with ELL students poses more far complex and challenging process than with native speakers of English since it should be used to meet at least six purposes (O'Malley and Pierce, 1996): screening and identification, placement, reclassification or exit, monitoring students' progress, program evaluation and accountability. In terms of authentic assessment, forms of assessment that reflect student learning, achievement, motivation and attitudes on instructionally-relevant classroom activities, we may encounter such examples as performance assessment, portfolios, and student self-assessment.

V. A MODEL OF ONE-MEETING THEME-BASED EXTENDED RESPONSE

The major barrier to implement critical thinking assessment is teachers' perception of impracticality and complexities of critical thinking assessment. Theoretically, the use of extended response essay is effective to improve students' critical thinking ability (Ennis, 1993). However, the real practices in a classroom context are considered as something that is too complex and impractical.

A. Preparation

Preparation is related to preparing every aspect needed in order to implement the theme-based extended response essay in classroom including the objectives, teaching materials, teaching procedures, and the instructions of the extended response essay. Teaching objectives in this study is that students will be able to produce short extended response essay reacting theme given by the teacher critically through various perspectives. In addition, a worksheet consisting of mind map, observation sheet, and essay instructions is also developed (Appendix 1).

In implementing the extended response task, teachers should stimulate the students by giving them prompts before asking them to go further with the task. The prompts can be in the form of videos, pictures, news, or other relevant prompts that can lead the further to understand more about the theme. In this case, Ethnicity is chosen as the major theme of the essay. Since the study is going to be conducted in Probolinggo, East Java, the prompt that is going to be given to the students is the video of Tengger tribe which lives in National Bromo Semeru Mountain Park.

The video of Tengger ethnic is selected as appropriate material of this study due to some reasons:

- It can provide various sub-themes for students to observe;
- The content is familiar since the ethnic is well-known as one of ethnic from Probolinggo;
- The visualization is clear to give general illustration about the content of the video; though they do not get the whole words in the video, they still can understand the content.

B. Procedures of Teaching

The video is going to be played in front of the class. Then, students have to observe the video and filling the mind-map of interesting and unique things of Tengger tribe that they have seen in the video. Next, the result of mind-mapping activity is going to be discussed.

The previous discussion will lead students to find some interesting aspects and uniqueness of Tengger tribe. These aspects then will be set as the sub theme of the task. Students are going to be grouped into several groups. Each group should focus on one subtheme. They have to observe and gather information of their subtheme from books, internet, and other relevant literature. The result of the observations is written in the observation sheet.

Next, after having group observation and reviewing the literature, students are asked to individually write one paper short essay reporting their observations.

C. Assessment

Bloom taxonomy higher order thinking skills (1956) consist of analysis, evaluation, and creating. Theme-based extended response can be used to develop all these critical thinking skills. Analysis skill defines as separating components or materials so that its organizational structure can be seen (Bloom, 1956). In this case, students try to separate the whole concept of Tengger tribe based on various subthemes and do in-depth observation about the subthemes. Evaluation defines as making judgment about ideas (Bloom, 1956). Students have to evaluate the information and data of their observation about Tengger ethnic. They have to decide whether the information is suitable or not. Next, creating skill defines as the way to express and communicate personal point of view (Bloom, 1956). Here, the creating stages is implemented through writing extended response essay based on their observations. Facione (2011) also explains the terms of six cognitive skills of critical thinking: interpretation, analysis, evaluation, inference, explanation, and self-regulation. Although the theme-based extended response essay can be implemented in order to develop all of critical thinking skills, in this study, the cognitive skills that are implemented are interpretation, inference, and explanation. Facione (2011) defines interpretation as comprehending and expressing meaning based on experiences, situations, data, and beliefs. The interpretation skill is used when students had to observe the literature for their essay. Then, inference skill defines as the ability to consider relevant information and determining the consequences of using the data, statements, and other form of representation (Facione, 2011). Next, explanation stage is used while students had to explain their ideas in the form of written essay.

Based on the stages of higher order thinking skills above, a writing rubric is developed in order to measure how critical the extended response short essay is (Appendix 2).

VI. CONCLUSION

Critical thinking assessment is crucial to be implemented in the classroom context in order to develop not only students' competences, but also their critical thinking skill. Theme-based extended response essay is chosen as the tool of critical thinking assessment in English foreign learning classroom since extended response essay enable the development of critical thinking skills. In addition, the various themes provided in this essay enable students to think one particular thing through various perspectives. Through critical thinking assessment, the students are expected to be able to improve their higher order thinking skill.

REFERENCES

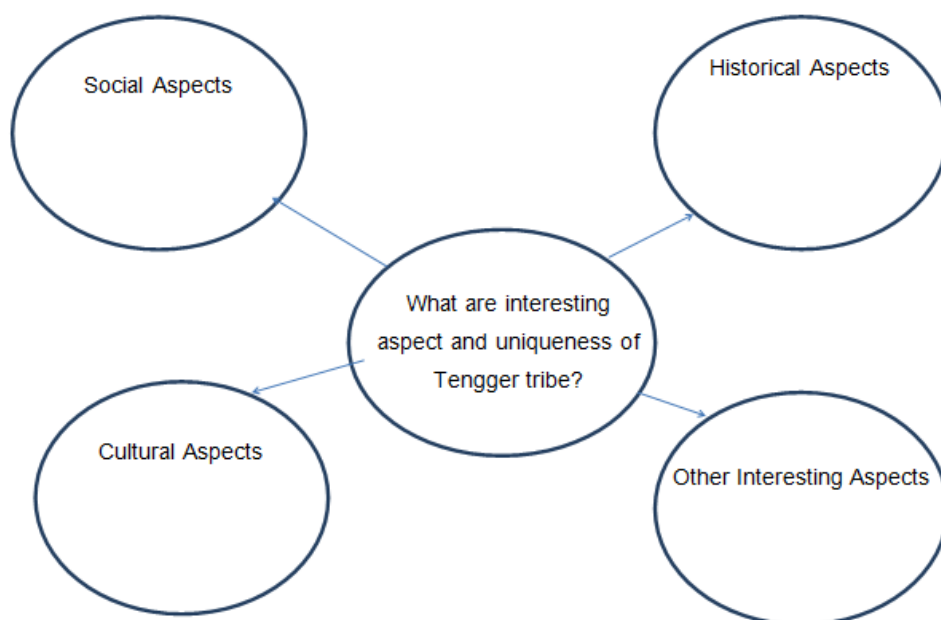
- [1] Bloom, B. 1956. "Bloom's Taxonomy in Learning Domains". Online: <http://www.nwlink.com/~donclark/hrd/bloom.html>.
- [2] Ennis, R.H. 1993. "Critical Thinking Assessment". Theory into Practice Volume 32, Number 3, Summer 1993.

- [3] Facione, P.A. 2011. "Critical Thinking: What It is and Why It Counts". Millbrae. The California Academic Press.
- [4] O'malley, J.M. & Pierce, L.V. 1996. "Authentic Assessment for English Language Learners". Pearson Education.
- [5] Smyth, J. (2000). "Reclaiming Social Capital through Critical Teaching". *The Elementary School Journal*, vol. 100 no. 5, pp. 491–511.
- [6] Stiggins, R. C. 1991. "Passing the Challenges of a New Era of Educational Assessment". *Applied Measurements in Education* vol 4 (4): 263-272.
- [7] Youtube Channel. "Tengger Tribe in Bromo Mountain". Online <https://www.youtube.com/watch?v=Qd9UE5LsWI8>.

Appendix 1. One Meeting Theme-based Extended Response Worksheet

Part I. Mind Map

Watch the video titled "Tengger Tribe" carefully. Write what are the interesting aspects and uniqueness of Tengger tribe that you see in the video in the mind map.



Part II. Observation Sheet

Write information about interesting aspect of Tengger ethnic as many as possible. You have to do mini research by reviewing some literature in the internet, books, and other relevant sources.

Group:

Subtheme:

No	Findings
1.	
2.	
3.	
4.	
5.	
Etc.	

Part III. Use the answers of the previous observation about Tengger tribe to write one paper essay reflecting your point of view analytically.

Appendix 2. Writing Rubric

Criteria	3	2	1
Do students involve critical analysis, interpretation, inference, and point of view towards the subtheme that they explore?	Perfectly critically displayed and presented in the essay.	Only 80% critical in exploring the topic	Less than 60% critical in exploring the topic.
Do the students write the essay with good coherency?	Perfect coherence of the paragraph	80% coherence of the paragraph	60% coherence of the paragraph
Is the essay written in comprehensive organization of paragraphs?	Perfect organization of the paragraph	80% organization of the paragraph	60% organization of the paragraph

Application of Instructional Model of Daily Assessment for Improvement of Processes Quality and Instructional Outcomes

Benidiktus Tanujaya

Department of Mathematics Education

University of Papua, Manokwari, West Papua, Indonesia

benidiktus.0903@gmail.com

Abstract—This research is conducted to improve the quality of process and outcome of instructional using daily assessment model. Daily Assessment is an Instructional Model that applies the principles of observation and assessment every day, especially on each of the instructional process. Some of the approaches taken in the application of this model are the use of the collaborative method, group and individual tasks, the group and individual presentation, as well as discussion and quiz conducted on every learning activity. The results showed that students responded very well to the model used that led to significant increase of students' learning activity. The increasing performance of students was shown by the students' test results

Keywords: *Daily assessment, Learning model, Learning process*

I. INTRODUCTION

Instructional is a process of interaction between students and teachers in a learning environment where teachers and students exchange information. There are a variety of instructional objectives, but basically, instructional objective is the behavioral changes that are expected to occur, owned and controlled by learners after participating in instructional activities. In order to achieve the maximum result of instructional objectives, then the quality of the instructional process needs to be improved.

Improvement of instructional quality can be done in various ways and approaches, using some models, methods and strategies of instructional. Improving the quality of teaching and learning can also be done by using the assessment. Assessment is vital to the instructional process. Assessment is a technique and instrument that is essential in the instructional process. Traditionally, assessment is used to measure how much students have learned a particular course point in a specific range of time. Assessment is also used to improve the quality of instructional process. Assessment is an approach designed to help lecturers find out what students are learning in the classroom and how well they are learning it. The assessment should be related to the instructional outcomes established for the instructional.

There are two types of assessment, summative and formative assessment. Summative assessment are used to measured what student have learnt at the end of a unit, to promote students, to ensure they have met required standards on the way to earning certification for school completion or to enter certain occupations, or as method for selecting students for entry into further education. On the other hand, formative assessment refers to frequent, interactive assessment of student progress and understanding to identify learning needs and adjust teaching appropriate [1]. In short, the primary purpose of formative assessment is the improvement of instructional outcomes. However, it cannot be implemented properly yet.

The problem is how to conduct formative assessment effectively in order to improve the quality of instructional process and instructional outcomes. The assessment should be useful for both teachers and students. Benefits for teachers, among others, is that it is not difficult to assess, as for students, the assessment can be directly used to increase their learning performance. Upon using the results of the assessment, the students are expected to know what is known and what is unknown, the students know what to do and what not to do. In short, the students will learn their strengths and weaknesses in instructional process. In addition, students will also learn what action should and needs to do to increase the quality of instruction. In other words, students should use the feedback on assessment to improve their learning and the quality of the instructional process

Assessment, especially formative assessment is essentially giving feedback both to the teacher and to the student regarding present understanding and skill development in order to determine the way forward. Feedback on assessment is an important part of the instructional process. However, both students and lecturers frequently show disappointment and frustration in relation to the conduct of the feedback process. Student complain that feedback on assessment is unhelpful, unclear, or sometimes even upsetting. Furthermore, students show that they are not given guidance as to how to use feedback to improve subsequence performance. Even worse, students also notice that the feedback is handed back to the students either too late or less relevance to their needs. On the other hand, lecturers frequently make comment that students are not interested in the feedback comments and are only concerned with the grade. Additionally, lecturers express frustration that student happen to not incorporate feedback advice into subsequent task [2].

Assessment feedback need to be given as soon as possible upon the completion of the learning task. Student also need to see that feed-forward comments can be incorporated into subsequent performance and overall influence the quality of their learning in positive ways [2]. Furthermore, temporarily withholding feedback is needed to allow the students to internalize and process the demands of the task [3].

Therefore, teachers need an [instructional model](#) so that they can make the assessment as soon as possible without having to wait until the middle or end of the semester. At the same time students/teachers can find out his weaknesses as soon as possible. Thus, students can undertake the necessary efforts so that the learning objectives can be achieved.

Thus, there are some questions that need to be answered. How to improve the quality of learning and teaching by using the assessment? Is it true that the quality of learning can be enhanced by using assessment? How can the quality of learning and teaching be improved by using the daily assessment? When should the daily assessment should be used? How can daily assessment work more effectively to improve instructional outcomes? In order to answer these questions, it is necessary to conduct a study to prove that the instructional model of daily assessment can be used to enhance the process and the outcome of instructional.

II. RESEARCH METHOD

This research was conducted in the University of Papua in Manokwari West Papua, It was participated by approximately 40 students majoring in Mathematics Education. The students studied are from three classes that researchers taught during the semester.

This research was carried out by using action research method. In this type of research, there is a reflexive process in which lecturers systematically study the problems in order to guide, correct, and evaluate the decisions and actions regarding to the improvement of instructional. There are several procedures [in running](#) this research, as follow:

- **Reconnainssance.** This stage had been done before the implementation of the lectures because of the character and abilities of each student lecture participants already well known by researchers. The students also have already had many years of learning in researcher's classes. At this stage, the researcher also did a series of observations of the characters of the course materials to be adapted in the assessment.
- **Application.** The second stage is the application stage of daily assessment model. Daily Assessment is a Learning Model that applies the principles of observation and assessment in daily basis, especially on each of the learning process. Some of the approaches taken in the application of this model are the use of the collaborative method, group and individual tasks, the group and individual presentation, as well as discussion on every learning activity.
- **Observation, data collection, and reflection.** In this stage, the reseracher did a series of observations to collect the data for the purposes of reflections. The reflections were done by the students participants and the teachers participants. This activity was meant to give both sides.

Below are several steps done by the researcher in order to check students' understanding;

Summaries, *Students demonstated what they had heard or read, and derived personal meaning from their learning experiences.*

Lists, Charts, and Graphic, *Students organize information, made connections, and noted relationships through the use of various graphics.*

Group Activities Students have the opportunity to communicate with others as they develop and demonstrate their understanding of concepts.

- **Follow-up actions.** At this stage, any information obtained previously, is occupied to formulate action in the future activity. The data obtained in this study were analyzed with students as the class continues. Students are requested to figure out its weaknesses through self-evaluation, the assistance of the group mate, and a peer from the other groups. Researchers in this case merely acts as a facilitator and motivator.

In addition to performing the daily assessment, the researchers also used the test instrument and the non-test instrument at the beginning, middle, and end of the course. The non-test instrument was used to measure the various activities of students in instructional, while the test instrument was to measure students' ability to understand the learning materials. Some of the students' activities were observe were: the ability to ask, the ability to answer, the ability to propose ideas, activity, and discipline. In addition, the performance of the students in understanding the learning materials to be measured are knowledge, comprehension, application, analysis, and synthesis.

III. RESEARCH FINDING AND DISCUSSIONS

A. Instructional Activity

Instructional activity means student' activity during the lectures that have a direct relationship with their instructional achievement. Results of research on instructional activities shown by students are presented in Table 1.

TABLE 1. DEVELOPMENT OF STUDENT INSTRUCTIONAL ACTIVITIES ON SOME INDICATORS

Indicator	Number of Student (%)		
	<i>Before</i>	<i>mid</i>	<i>after</i>
Asking	32.50	45.00	85.00
Answer	40.00	52.50	70.00
Activeness	50.00	65.00	70.00
discipline	62.50	70.00	80.00
Asking Idea	25.00	40.00	60.00

Table 1 shows that there are significant changes in several indicators during the research. Students tend to have better learning activity at the end of the study compared to the early and mid-lecture. The study showed that the highest percentage in instructional activities of students was in their ability in asking questions.

Questioning skills is an ability that is very important in learning. Asking questions is the desire to find information that is not yet known. In general, education experts [believe](#) that students who have the good ability to ask tend to have good learning performance as well.

Students performance improving in several indicators studied, particularly the ability to deliver question due to use of the [instructional model](#) of daily assessment. There are several types of assessment used in this study. By varying the type of assessment, according to [4], lecturers can get a more accurate picture of what students know and understand, obtaining a "multiple-measure assessment 'window' into student understanding". Using at least one formative assessment daily enables lecturers to evaluate and assess the quality of the learning that is taking place in your classroom and answer these driving questions: How is this student evolving as a learner? What can I do to assist this learner on his path to mastery?

B. Instructional Outcome

The instructional outcomes are the ability shown by students after studying a particular subject. Instructional outcomes are learning outcomes that are observable, measurable, and assessable statements of the end product of student learning including knowledge, skills, competencies, and attitudes. Students performance on instructional outcomes observed in this study are presented in Table 2 below.

TABLE 2. DEVELOPMENT OF STUDENT'S PERFORMANCE ON SOME INDICATOR

Performance	Percentage (%)		
	<i>Before</i>	<i>mid</i>	<i>after</i>
Knowledge	52.50	70.00	87.25
Comprehension	60.00	62.50	77.25
Application	50.00	67.25	70.00
Analysis	42.50	50.00	60.00
Synthesis	35.00	55.00	60.00

Table 2 shows that a significant increase in student performance in all aspects observed. The success of the students in the activities and performance in the instructional benefited by using this model. By making the daily assessment led to students will learn their weaknesses, what should be corrected, what should be improved, what should to avoid, and what they should do to improve learning achievement.

The use of the assessment as an instrument for the evaluation has been done. However, students generally do not use the results of the assessment properly. Some education experts stated that: It is also possible that student do not pay attention to comments because they don't make sense to them [5] or that they do not understand the purpose of the feedback process. This is accentuated when feedback is delivered solely by the lecturers and is often associated with students as the marking of what is right and wrong. Many lecturers may also tend to focus on the correctional rather than the instructional aspects of feedback [3].

C. Correlation between Performance and Activities of Students

Correlation is a statistical value that indicates the closeness and direction of the relationship between the two variables. Correlation between performance and activities of student is a value that indicates the relationship between the two variables. Correlation between performance and activities of student observed in this study are presented in Table 3.

TABLE 3. CORRELATION OF SOME PERFORMANCE AND ACTIVITIES OF STUDENT

Performance	Instructional Activities				
	Asking	Answer	Activeness	discipline	Asking Idea
Knowledge	0.76	0.67	0.63	0.78	0.70
Comprehension	0.66	0.67	0.60	0.70	0.70
Application	0.65	0.66	0.66	0.70	0.62
Analysis	0.86	0.77	0.80	0.86	0.78
Synthesis	0.70	0.60	0.65	0.56	0.64

Table 3 shows that there are significant correlations between students' performance and activities. A very close relationship occurs between the analysis and all components of students' activities. This shows that the better the learning activity, the better the performance achieved by the students, vice versa.

This results from the implementation of daily assessment model, in which students and lecturers work together. This instructional model does not only belong to the lecturers, but also to the students as this model is an integral part of instructional process. According to [6], when teacher' classroom assessments become an integral part of the instructional process and a central ingredient in their effort to help students learn, the benefits of assessment for both students and teachers will be boundless.

In addition, during the implementation of daily assessment model where students perform self-evaluation, researchers merely play a role as the observer. This helps lecturers to carry out more qualified learning activities. In this parts, the students deliver carefully listen to feedback, deliver thoughtful questions, and give reflective responses.

IV. CONCLUSIONS AND SUGGESTIONS

Based on the results of research and discussion above, it can be concluded that the *instructional model* of daily assessment can improve the *instructional* process and *instructional* outcomes of mathematics education students UNIPA. Students are becoming increasingly active in a variety of instructional activities. The instructional model also provides an optimal instructional outcomes.

Based on the research results, there is an alternative way to increase instructional process and instructional outcome, it can be suggested on the following steps:

1. Provide the student to understand the learning objective.
2. Learn the characteristics of students and the characteristics of the subject matter.
3. Choose the appropriate form of assessment for each instructional activity.
4. Teach students to assess themselves.

Several steps done by the researcher in order to check students' understanding:

1. Summaries, Students demonstrate what they have heard or read, derive personal meaning from their learning experiences.
2. Lists, Charts, and Graphic, Students organize information, make connections, and note relationships through the use of various graphic.
3. Group Activities, Students have the opportunity to communicate with others as they develop and demonstrate their understanding of concepts.

ACKNOWLEDGMENT

The author would like to thank everyone contributed in this research, especially the students of Mathematics Education UNIPA, who have given so many ideas in instructional activities for this research. Thanks also to my lovely-wife Jeinne and my sweet-daughters Aurelia, for all the support.

REFERENCES

- [1] OEDC. Assessment for Learning: Formative Assessment. International Conference, "Learning in the 21st Century: Research, Innovation, and Policy" Center for Education Research and Innovation, 2008.
- [2] Spiller, Dorothy. Assessment: Feedback to Promote Student Learning. Teaching Development Unit, New Zealand, February 2009.
- [3] Hattie, J and H. Timperley, The Power of Feedback, *Review of Educational Research*. 2007, 77, pp.81–112.
- [4] Dodge Judith. What Are Formative Assessments and Why Should We Use Them?
- [5] <http://www.scholastic.com/teachers/article/what-are-formative-assessments-and-why-should-we-use-them>
- [6] Duncan N. Feed-forward: Improving Students' Use of Tutor Comments, *Assessment and Evaluation in Higher Education*. 2007, Vol. 32 (3), pp. 271-283.
- [7] Guskey, T. R. How Classroom Assessments Improve Learning. *Educational Leadership*. 2003. 60:5, pp.6-11.

ASSESSING STUDENT'S PRAGMATICS' KNOWLEDGE AT ISLAMIC UNIVERSITY OF RIAU

Betty Sailun

(English Study Program, Islamic University of Riau)

Betty.sailun@edu.uir.ac.id

Joko Ariyanto

(English Study Program, Islamic University of Riau)

ariyanto87@edu.uir.ac.id

ABSTRACT— Pragmatics is defined as appropriate use of language either to comprehend ideas or to interact in social situations effectively. Pragmatic competence, which is processed in the right hemisphere, comprises a number of interrelated skills that manifest in a range of adaptive behaviors. Due to the widespread influence of language in communication, studying pragmatic profiles, by developing appropriate pragmatics assessment, is of importance. This study explored ways to assess students of English department Islamic University of Riau pragmatics competence. In this study the researcher used Written Discourse Completion Tasks (WDCT), Multiple-Choice Discourse Completion Tasks (MDCT), and Discourse Self-Assessment Talks (DSAT). Pragmatics components that were analyzed are conversational skills in social interaction and Speech Acts. The participants of the research were the fifth semester students of English department in Islamic University of Riau within the subject of Pragmatics. The findings reconfirmed the importance of teaching pragmatic knowledge to English learners in classrooms. The pedagogical implications and applications for English language teachers are also discussed. The paper concludes with the suggestion that English teachers should teach pragmatic knowledge in class and include inter language pragmatic knowledge in large-scale tests.

Keywords: *Assessment, Pragmatics Knowledge, Islamic University of Riau*

I. INTRODUCTION

Assessment As a domain within L2 studies, pragmatics is usually referred to as interlanguage pragmatics (ILP), as analogy with interlanguage grammar, interlanguage phonology, and interlanguage lexicon (Kasper & Rose, 2002). ILP is a second-generation hybrid (Kasper & Blum-Kulka, 1993). It belongs to two different disciplines, both of which are interdisciplinary. On one hand, as a branch of second language acquisition research, two sections within the wider domain of ILP are distinguished. As the study of L2 use, ILP examines how nonnative speakers (NNSs) comprehend and produce action in a target language. As the study of L2 learning, ILP investigates how L2 learners develop the ability to understand and perform action in a target language (Kasper & Rose, 2002). On the other hand, as a subset of pragmatics, ILP is a sociolinguistic, psycholinguistic, or simply linguistic enterprise, depending on how one defines the scope of pragmatics (Kasper & Blum-Kulka, 1993).

Kasper and Blum-Kulka (1993, p. 3) define ILP as “the study of nonnative speaker’s use and acquisition of linguistic action patterns in a second language”. They also offer a broader definition of ILP. They argue that “tying interlanguage pragmatics to NNSs, or language learners may narrow its scope too restrictively” (p. 3), and include into ILP “the study of intercultural styles brought about through language contact, the conditions for their emergence and change, the relationship to their substrata, and their communicative effectiveness” (p. 4). But most ILP studies focus on the narrow definition. Kasper’s later definition of ILP also focuses on the narrow sense. Kasper (1998, p. 184) defines ILP as the study of nonnative speakers’ comprehension, production, and acquisition of linguistic action in L2, or put briefly, ILP investigates how to do things with words in a second language.

In this study, interlanguage pragmatic knowledge is defined, according to Kasper (1998) and Rose (1997), as the nonnative speaker's knowledge of a pragmatic system and knowledge of its appropriate use. Since the idea of interlanguage pragmatics was introduced into language education, it has received more and more attention in language courses, such as through the notionally-based syllabus (Cohen & Olshtain, 1981). Studies have been done to investigate the relationship between language education and interlanguage pragmatic development, for example, whether grammatical development guarantees a corresponding level of pragmatic development. The results of these studies differ. Some studies (e.g. Hill, 1997; Roever, 2005; Yamashita, 1996) showed that high language proficiency participants had better performance in tests of pragmatics than low language proficiency participants in an English as second language context. On the other hand, other studies (e.g. Bardovi-Harlig & Hartford, 1991, 1993; Omar, 1991; Takahashi & Beebe, 1987) showed disparities between learners' grammatical development and pragmatic development. They reported that even learners who exhibit high levels of grammatical competence may exhibit a wide range of pragmatic competence when compared with native speakers in conversations and elicited conditions (Bardovi-Harlig & Doernyei, 1998). He and Yan (1986) investigated the pragmatic failure by Chinese learners of English as a foreign language (EFL) and found that the learners' pragmatic development was not proportional to their grammatical development.

Meanwhile, some studies have been done to investigate the teachability of pragmatic knowledge in classrooms and some (e.g. Bardovi-Harlig, 2001; Fukuya, Reeve, Gisi, & Christianson, 1998; Golato, 2003; Matsuda, 1999; Rose & Kasper, 2001) have shown that interlanguage pragmatic knowledge is teachable. The necessity and importance of teaching pragmatics have also been recognized (Eslami-Rasekh, 2005; Rose & Kasper, 2001), but still language teachers hesitate to teach pragmatics in their classrooms. Thomas (1983) notes that for the language teachers the descriptions offered by theoretical pragmaticists are inadequate. Matsuda (1999) lists two reasons for this reluctance in pragmatics teaching. First, teaching pragmatics is a difficult and sensitive issue due to the high degree of *Assessing EFL learners' interlanguage pragmatic knowledge* 'face threat' it often involves and, second, the number of available pedagogical resources is limited. But the reluctance should also be attributed to the lack of some valid methods for testing interlanguage pragmatic knowledge. More studies need to be done to validate methods for pragmatics assessment. The present study aims to investigate ways to assess EFL learners' pragmatic competence by addressing two questions:

- 1) Are the test methods used in this study reliable and valid?
- 2) Do learners of different EFL proficiency levels perform differently in pragmatics tests?

II. WAYS TO ASSESS INTERLANGUAGE PRAGMATIC KNOWLEDGE

Oller (1979) first introduced the notion of a pragmatic proficiency test and set two constraints for this kind of test. First, processing of language by examinees on pragmatic tests must be constrained temporally and sequentially in a way consistent with the real world occurrences of the language forms that happen to comprise test materials or speech in testing situations. This constraint could imply, for example, that encountering sentences on a reading comprehension test would require that an examinee process such sentences as meaningful sentences, rather than as just strings of words with no communicative intent. Second, such tests must use language in a way resembling natural occurrences of language outside testing contexts or formal language testing environments.

The meaning of language understood or produced in pragmatic tests must link somehow to a meaningful extralinguistic context familiar to the proficiency examinee. Oller stressed the naturalness of such a test. These naturalness criteria, however, seem problematic, because they do not adequately address the artificiality of testing contexts in and of themselves, and how such artificiality constrains language use (Duran, 1984). This issue was better addressed by Clark (1978) through the notion of direct versus indirect tests of language proficiency. Clark suggested that a 'direct' test should be based on approximating, to the greatest extent possible within the necessary constraints of testing time and facilities, the specific situations in which the proficiency is called upon in real life. Clark indicated that direct proficiency tests should model everyday language use situations, but he also acknowledged that testing contexts could only approximate the real world. Unfortunately, the field of language testing does not seem to offer much research in this respect. Not many tests to assess learners' pragmatic proficiency have been produced, though pragmatic knowledge is an indispensable part of language proficiency as defined by Bachman (1990).

One of the reasons why such measures have not been readily available is that developing a measure of pragmatic competence in an EFL context is not an easy task. So far, researchers have investigated at least six types of methods for interlanguage pragmatic assessment, i.e., the Written Discourse Completion Tasks (WDCT), Multiple-Choice Discourse Completion Tasks (MDCT), Oral Discourse Completion Tasks (ODCT), Discourse Role Play Talks (DRPT), Discourse Self-Assessment Talks (DSAT), and Role-Play self-assessments (RPSA). A summary of the practical characteristics of the six types of tests is given in Brown (2001a). The six measures are reviewed in detail in Yamashita (1996) and Yoshitake-Strain (1997). Brown and Hudson (1998) classified language assessment into three broad categories: selected-response assessments, constructed-response assessments, and personal-response assessments. For the sake of representativeness, in this study one test method from each of these three categories was selected: WDCT from the constructed-response type, MDCT from the selected-response type, and DSAT from the personal-response type. The following is an introduction to the forms of the three test methods used in this study.

A. Written discourse completion test

WDCTs are written questionnaires including a number of brief situational descriptions, followed by a short dialogue with an empty slot for the speech act under study. Participants are asked to provide a response that they think is appropriate in the given context: At the professor's office

A *student* has borrowed a book from her teacher, which she promised to return today. When meeting her teacher, however, she realizes that she forgot to bring it along.

Teacher: *Miriam, I hope you brought the book I lent you.*

Miriam:

Teacher: *OK, but please remember it next week.*

(Blum-Kulka & Olshtain, 1984, p. 198)

WDCTs have evolved gradually over the past twenty years into several different modified versions which vary mainly according to the presentation forms, that is, written or oral, and existence of rejoinder. WDCTs can include a rejoinder, as in the following example from Johnston, Kasper, and Ross (1998, p. 175): *Your term paper is due, but you haven't finished yet. You want to ask your professor for an extension.*

You:

Professor: *I'm sorry, but I never allow extension.* Or they may involve only the specification of the situation with no rejoinder, as this example from Eisenstein, Bodman, and Carpenter (1996, p.102) shows:

Two people who are friends are walking toward each other. They are both in a hurry to keep appointments. They see each other and say:

In this study, this type of WDCT with no rejoinders was adopted.

B. Multiple-choice discourse completion test

MDCTs consist of test items where the test taker is required to choose the correct response (the key) from the several given options. Most commonly, multiple-choice items include an instruction to the test taker and a stem (typically either a phrase or sentence to be completed, or a question). The key and several *Assessing EFL learners' interlanguage pragmatic knowledge* distractors then follow in random order (Davies et al., 1999). Following is a sample MDCT item:

You are a student. You forgot to do the assignment for your Human Resources course. When your teacher whom you have known for some years asks for your assignment, you apologize to your teacher.

A. *I'm sorry, but I forgot the deadline for the assignment. Can I bring it to you at the end of the day?*

B. *Pardon me, sir, I forgot about that. Shall I do the assignment at once? So sorry! It's my fault!*

C. *I've completed my assignment but forgot to bring it with me. I'll hand it in tomorrow.*

C. Discourse self-assessment test

On the DSAT, instructions are first given, followed by exponents of the functions. The participants, after reading each situation, were asked to give an overall rating of their intended

performance on a five-point scale. The following is an example of the self-assessment from Hudson, Detmer & Brown (1995, p.192).

Situation: You and a few of your co-workers are working on a special project. You are at a meeting in the office of the project leader. As you are reaching for your briefcase you accidentally knock over the project leader's umbrella which was leaning against the desk. Rating: I think what I would say in this situation would be very unsatisfactory 1 – 2 – 3 – 4 – 5 completely appropriate

III. METHOD

A. Participants

69 Islamic University of Riau EFL learners participated in this study in the final data collection stage. All of them were students from fifth semester, of which the third-year students. high language proficiency group (hereinafter called high level group) and the first-year students (N=58) were taken as the low language proficiency group (hereinafter called low level group). The ages of the participants ranged from 16 to 21 for the low level group (mean=18.79) and 19 to 24 for the high level group (mean=21.06), with an overall average of 19.61. The first-year students had studied English for about seven years, and the third-year students about 10 years. To further validate the proficiency levels of the students, a proficiency test (TOEFL, with permission granted by the Educational Testing Service) was administered to both groups. Table 1 shows the statistical analyses of the proficiency test for the two groups. From this table, we can see the two groups differ significantly in all sections of the test: listening ($t=6.04$, $p<.01$), structure ($t=4.30$, $p<.01$), reading ($t=6.38$, $p<.01$), and total ($t=8.34$, $p<.01$). This shows that the two groups were significantly different in terms of their English proficiency levels.

In addition, a questionnaire was administered to gather more background information about the participants. The questionnaire consisted of five questions about gender, age, experience in English-speaking countries, pragmatic knowledge taught in class, and self-assessment of the ability to use English. None of the participants reported any experience of having stayed in an English-speaking country.

Assessing EFL learners' interlanguage pragmatic knowledge designed in which participants were given a sheet of paper illustrating a request. A brief training session was conducted before the students began to answer the questionnaire. The questionnaire was first explained to the students. Then, some of the students were asked to provide an example of the desired speech act. A brief follow-up discussion was held so that the students knew what they were supposed to do. Altogether 69 students were asked to complete the exemplar generation questionnaire. Each student was asked to write the 10 most recently occurring events which contained the speech act of requesting. All 69 students returned their questionnaire and most wrote 10 situations. It was found that many of the nearly 300 situations generated by the students were similar; consequently only 57 situations were selected. All these situation scenarios were rewritten with their original meaning basically unchanged. Reference was also made to some existing scenarios from the literature.

B. Likelihood investigation

Next, a likelihood investigation was conducted for the 57 situations. This questionnaire asked the respondents to indicate on a scale of 1 to 5 the likelihood that the situations would occur in their daily life. The likelihood investigation questionnaire was written in Indonesian. The questionnaire was sent to 15 students. The scales selected by the respondents were averaged for each situation. The 30 situations which got the highest mean scores were selected to form the metapragmatic assessment questionnaire.

C. Metapragmatic assessment

The 30 situations were reviewed. Priority was given to those situations with different combinations of features. As a result, only 24 situations were used in the metapragmatic assessment questionnaire. The questionnaire, with detailed instructions and specific examples, asked the respondents to indicate the imposition of the situation, the social distance (familiarity) between the speaker and hearer, and the power relationship (status) of the hearer and speaker (who has higher status if not equal). The questionnaire for the CUSs was written in Indonesian, while that for the English native speakers was in English. The two versions were generated

through back translation (Brislin, Lonner, & Thorndike, 1973; Fowler, 1993; Kasper & Rose, 2002; Sinaiko & Brislin, 1973).

Next, a new MDCT questionnaire was designed with these 12 situations and their corresponding options. This MDCT questionnaire was given to 5 ENSs who were asked to select the most appropriate response for each situation. Their responses showed unanimous agreement on the keys for 12 situations. In the next step, 31 CUSs were invited to answer the draft MDCT questionnaire. The questionnaire was also used for the think-aloud protocol with two participants. The verbal protocols were recorded, and analyses using item response theory (IRT) were conducted for the paper-and-pencil questionnaire. The think-aloud data were also analyzed to see if the students were able to determine the intended speech act, and to examine how the distractors functioned. Corresponding revisions were made according to the results of the IRT analyses. *Assessing EFL learners' interlanguage pragmatic knowledge* The analyses of the verbal protocols showed that all the items tested the right speech act they were intended to do.

Finally, three test papers using these three test methods (i.e., DSAT, WDCT, and MDCT) were generated. All three test papers contained the same 12 situation scenarios, but with different test methods.

D. Administrative procedures

The tests were administered to 89 university students in three sessions. A proficiency test (TOEFL) was first administered. Two weeks later, the other tests were administered to the two groups (i.e., the high-level group and the low-level group) in turn in a classroom. The DSAT was administered first, followed by the WDCT and then the MDCT. The whole test session took about two hours.

E. Scoring

To avoid any effect on ratings due to poor handwriting, the answers to the WDCT test paper given by the test takers were entered into the computer without any changes. The typewritten scripts were ordered alphabetically according to the test takers' surnames and then presented to two ENS raters using the rubrics developed by Hudson, Detmer, and Brown (1995). The raters were given clear directions as to how the test papers should be rated and had a preliminary training on the rating. The final scores of the WDCT test were the mean scores of the two raters. For the MDCT, one correct answer equaled to five points while a wrong answer got 0 point. For the DSAT, the test takers' self-rating was the final score.

IV. RESULTS

The WDCT yielded lower mean scores than the MDCT, and it was the DSAT that produced the highest mean scores. This shows that the WDCT test paper was more difficult than the MDCT test paper and that the participants overestimated their pragmatic ability to some degree.

A. Internal consistency reliability

The internal consistency (Cronbach's alpha) of each test was estimated. Then, the internal consistency reliability for ratings on the WDCT was examined. The correlation between the two raters was also computed to further show the interrater reliability. Table 3 displays the results of the estimates and the standard error of measurement for each test method. The internal consistency reliability estimates for all the test methods were basically satisfactory (all were above .86).

The internal consistency reliability estimates for the two raters were acceptable (.87 for Rater 1 and .75 for Rater 2). This suggests that there was a considerable amount of consistency in assigning scores to the examinees' performance, although there existed some disagreement between the two independent raters. The overall interrater reliability based on the Spearman-Brown Prophecy formula was .903. Meanwhile, the interrater reliability can also be estimated by examining how much the rater's scores correlate with each other (Yamashita, 1996). The correlation ($r=.82$, $p<.01$) indicated that the two raters were significantly correlated.

B. Correlational analyses

Correlational evidence was collected to examine the relationship among the items and test methods. Table 4 displays the correlations between different tests. The TOEFL test was not significantly correlated with other tests except the MDCT. The three pragmatics tests correlated

significantly with each other at $p < .01$. Statistical significance is a necessary precondition for a meaningful correlation, but it is not sufficient unto itself (Brown, 1996). So, coefficients of determination for these correlations were computed.

V. DISCUSSION AND CONCLUSION

Reliability and validity can be viewed as complementary aspects of validation process (Bachman, 1990). The Cronbach alpha reliability estimates for WDCT and DSAT were satisfactory at around .90, while that for MDCT was .86. This is in line with previous studies (Enochs & Yoshitake-Strain, 1999; Roever, 2005; Yamashita, 1996) which showed that WDCT and DSAT had high reliability. Nevertheless, it is noteworthy that the internal consistency reliability for the MDCT in this study was acceptably high at .86. This might be due to the procedures involved in the development of the test paper. The scenarios and options of the MDCT test paper in this study, instead of adopting established ones, were independently developed in several stages based on the Chinese context, including exemplar generation, likelihood investigation, metapragmatic assessment, and verbal protocol analysis. All the situations were closely related to the participants' life, and the distractors were generated by the participants. However, although the MDCT test paper developed in this study worked well for the Chinese context, it is not clear whether it would work equally well in other contexts, or with different participant groups. More research is needed.

The significant correlation between the MDCT and the TOEFL can be attributed to their ability in tackling such questions. *Assessing EFL learners' interlanguage pragmatic knowledge* The two groups of participants were significantly different ($p < .01$) in terms of their English language proficiency, whereas this study revealed that they were not significantly different in the tests of pragmatics (WDCT and DSAT). However, the two proficiency groups were significantly different at the .05 level on the MDCT. This difference might result from the effect of the test method. The results indicated that the participants of higher grammatical proficiency did not necessarily possess higher concomitant pragmatic competence. This differed from some previous studies (e.g. Hill, 1997; Roever, 2005; Yamashita, 1996) which showed that high language proficiency participants had better performance in tests of pragmatics than low language proficiency participants.

There are two possible reasons for this difference. First, the participants in those studies had different degrees of experience in living in an English-speaking country. They had direct exposure to the target culture. However, the students in this study had no such experience or direct exposure. They were exposed to the target culture only through the classroom. An investigation also demonstrated no significant difference between the two proficiency groups in terms of pragmatic knowledge teaching in class. Second, the participants in those studies were rather diverse and heterogeneous; the students in this study, however, comprised only university students who had a similar educational background. They were divided into two proficiency groups according to their scores on a TOEFL test conducted just before the data collection. In fact, the English proficiency of the low-level group was not low at all, though significantly lower than the high-level group. Exposure to the target language is shown to affect the development of EFL learners' interlanguage pragmatic knowledge. One way of remedying this lack of direct exposure to the target culture and society may be through teaching pragmatics. Also, the insignificant difference between the two proficiency groups indicated that pedagogical measure should be taken to enhance the development of the EFL learners' interlanguage pragmatic knowledge.

However, investigation in this study showed that teachers seldom, if ever, taught pragmatic knowledge in class. The development of pragmatic competence, according to Ellis (1994), depends on providing learners with sufficient and appropriate input. Input in the EFL classroom comes mainly through teacher talk and instructional materials (Hill, 1997). However, foreign language teaching in Indonesian universities is conducted mainly in a traditional way in the classroom, that is, teacher-centered teaching. Though even teacher-fronted classroom discourse offers some opportunities for pragmatic learning (Kasper, 1997), and communicative teaching is receiving more and more attention in Indonesia nowadays, both ways still seem problematic in the Chinese EFL classroom.

First, the majority of the EFL teachers in Chinese universities are non-native speakers of English; thus, they cannot draw on native speaker (NS) intuitions (Rose, 1994) and cannot serve as direct models for the students (Bardovi-Harlig & Hartford, 1996). Feeling the lack of NS intuition also makes EFL teachers reluctant to teach pragmatics in classroom.

Second, insufficient instructional materials impede the move towards teaching pragmatics by EFL teachers. Being EFL teachers have difficulty in determining the appropriate materials for teaching pragmatics. Although evidence of speech acts in textbooks is plentiful, as pointed out by Bardovi-Harlig and Hartford (1996), it has been given very little attention. Therefore, Rose (1994, p. 155) notes that "if pragmatic competence is to be dealt with successfully in EFL settings, methods and materials must be developed which do not assume or depend on the NS intuitions of the teacher."

Third, the lack of instructional methods also prevents EFL teachers from teaching pragmatic knowledge in class. Two types of activities have been proposed for pragmatic knowledge instruction: activities aiming at raising students' pragmatic awareness, and activities offering opportunities for communicative practice (Kasper, 1997). Rose (1994) suggests pragmatic consciousness-raising in EFL teaching and comments that if the learner's pragmatic consciousness is raised, he or she will more easily notice pragmatic features of the input and this may lead to the acquisition of pragmatic knowledge. He also points out that in order for EFL learners to benefit from this type of consciousness-raising they should be given ample supplies of authentic input, e.g. through videos and movies.

On the other hand, practicing EFL learners' pragmatic abilities, however, requires student-centered interaction. Activities which engage students in different social roles and speech events, such as role play, simulation, and drama, provide opportunities to practice the wide range of pragmatic and sociolinguistic abilities that the students need in interpersonal encounters outside the classroom (Kasper, 1997). Owing to the factors which limit the use of these two types of activities, more practical activities for pragmatics instruction are expected.

Fourth, if pragmatic knowledge is included in the teaching syllabus, it needs to be incorporated into tests. However, no established tests of this kind are available now. Though some studies (e.g. Hudson, Detmer, & Brown, 1995; Liu, 2006; Yamashita, 1996; Yoshitake-Strain, 1997; Roever, 2005) examined the possibilities of such tests, at this time, as pointed out by Hudson (2001, p. 297), the instruments should be used for research purposes only, and no examinee level decisions should be made in pedagogical settings. More research, especially validation studies, is necessary.

This study has some implications for ILP test development, too. A fundamental concern, according to Roever (2005), in constructing items for tests of pragmalinguistic knowledge is that they be representative of real-world language use, and not just based on test designers' intuition, which may or may not be an accurate reflection of reality. Ethnographic studies of real world language use and targeted elicitation of response plausibility in pilot study are suggested (Roever, 2005). Ethnographic field study is a useful procedure, but the extremely expensive endeavors it demands precludes it from being a widely adopted method. Inconsistency might be found between elicitation through and that through NNSs (Yamashita, 1996). No such inconsistency was detected for the scenarios generated for this study. This would suggest that a combination of elicitation through both NSs and NNSs is a better and more practical way to construct pragmalinguistic test items.

This study also has some implications for the large-scale proficiency tests practiced nowadays around the world, like TOEFL and IELTS. The traditional paper-and-pencil TOEFL was found not to correlate with pragmatics tests. Test takers' English proficiency as shown on the TOEFL scores does not seem to be consistent with their interlanguage pragmatic ability. This study also showed *Assessing EFL learners' interlanguage pragmatic knowledge* that students with high TOEFL scores do not seem to have correspondingly high interlanguage pragmatic ability. Therefore, it is no wonder that some students who have gained over 670 points in a traditional paper-and-pencil TOEFL test cannot communicate well in English.

ACKNOWLEDGEMENTS

The study has been supported by the Lembaga Penelitian Islamic University of Riau. I would like to thank Dr Hasbi Zaidi for his constructive advice and comments. Thanks also go to the editors for their valuable feedback about and editing of the earlier drafts of the article.

REFERENCES

- [1] Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.
- [2] Bardovi-Harlig, K. (2001). Evaluating the empirical evidence: Grounds for instruction in pragmatics. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 13-32). New York: Cambridge University Press.

- [3] Bardovi-Harlig, K. & Doernyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic vs. grammatical awareness in instructed L2 learning. *TESOL Quarterly*, 32, 233-259.
- [4] Bardovi-Harlig, K. & Hartford, B. (1991). Saying 'no' in English: Native and non-native rejections. In L. F. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning. Monograph Series 2* (pp. 41-58). Urbana, IL: University of Illinois.
- [5] Bardovi-Harlig, K. & Hartford, B. (1993). Learning the rules of academic talk: A longitudinal study of pragmatic change. *Studies of Second Language Acquisition*, 15, 279-304.
- [6] Bardovi-Harlig, K. & Hartford, B. (1996). Input in an institutional setting. *Studies of Second Language Acquisition*, 18, 171-188.
- [7] Blum-Kulka, S. & Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5(3), 197-213.
- [8] Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods*. New York: John Wiley.
- [9] Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- [10] Brown, J. D. (2001a). Pragmatics tests: Different purposes, different tests. In K. R. Rose & G. Kasper. (Eds.), *Pragmatics in language teaching*. New York: Cambridge University Press.
- [11] Brown, J. D. (2001b). What is an eigenvalue? Retrieved August 3, 2004, from http://www.jalt.org/test/bro_10.htm
- [12] Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- [13] Clark, J. D. (1978). Psychometric considerations in language proficiency testing. In B. Spolsky (Ed.), *Approaches to language testing*. Arlington: Center for Applied Linguistics.
- [14] Cohen, A. D. & Olshtain, E. (1981). Developing a measure of sociocultural competence: the case of apology. *Language Learning*, 31(1), 113-134.
- [15] Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- [16] Duran, R. P. (1984). Some implications of communicative competence research for integrative proficiency testing. In C. Rivera (Ed.), *Communicative competence approaches to language proficiency assessment: Research and application* (pp. 44-58). Clevedon: Multilingual Matters. **18 Liu Jianda**
- [17] Educational Testing Service (2005). *TOEFL iBT Tips*. Retrieved February 8, 2006, from http://upload.mcgill.ca/applying/TOEFL_Tips.pdf.
- [18] Eisenstein, M., Bodman, J., & Carpenter, M. (1996). Cross-cultural realization of greeting in American
- [19] English. In J. Neu (Ed.), *Speech acts across cultures* (pp. 89-107). Berlin: Mouton de Gruyter.
- [20] Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- [21] Enochs, K. & Yoshitake-Strain, S. (1999). Evaluating six measures of EFL learners' pragmatic competence. *JALT Journal*, 21(1), 29-50.
- [22] Eslami-Rasekh, Z. (2005). Raising the pragmatic awareness of language learners. *ELT Journal*, 59(3), 199-208.
- [23] Fowler, J. F., Jr. (1993). *Survey research methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- [24] Fukuya, Y., Reeve, M., Gisi, J., & Christianson, M. (1998). *Does focus on form work for teaching sociopragmatics?* Paper presented at the annual meeting of the International Conference on Pragmatics and Language Learning, Urbana, IL.
- [25] Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, 24(1), 90-121.
- [26] Groves, R. (1996). How do we know what we think they think is really what they think? In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 389-402). San Francisco: Jossey-Bass.
- [27] Gyagenda, I. S., & Engelhard, G., Jr. (1998). *Applying the Rasch model to explore rater influences on the assessed quality of students' writing ability*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.
- [28] He, Z. & Yan, Z. (1986). Pragmatic failure by Chinese EFL learners. *Foreign Language Teaching and Research*, 3, 52-57.
- [29] Hill, T. (1997). *The development of pragmatic competence in an EFL context*. Unpublished doctoral dissertation, Temple University, Tokyo.
- [30] Hudson, T. (2001). Indicators for pragmatic instruction: some quantitative tools. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 283-300). New York: Cambridge University Press.
- [31] Hudson, T., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics*. Honolulu: Second Language Teaching & Curriculum Center University of Hawai'i at Manoa.
- [32] Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing Prototypic Measures of Cross-cultural Pragmatics*.
- [33] Honolulu: Second Language Teaching and Curriculum Center, University of Hawaii at Manoa.
- [34] Johnston, B., Kasper, G., & Ross, S. (1998). Effect of rejoinders in production questionnaires. *Applied Linguistics*, 19(2), 157-182.
- [35] Kasper, G. (1997). Can pragmatic competence be taught? Retrieved August 10, 1999, from <http://www.lll.hawaii.edu/nflrc/NetWorks/NW6/default.html>
- [36] Kasper, G. (1998). Interlanguage pragmatics. In H. Byrnes (Ed.), *Learning foreign and second languages: Perspectives in research and scholarship* (pp. 183-208). New York: The Modern Language Association of America.
- [37] Kasper, G., & Blum-Kulka, S. (1993). Interlanguage pragmatics: An introduction. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 3-17). Oxford: Oxford University Press.

- [38] Kasper, G. & Rose, K. R. (2002). *Pragmatic development in a second language*. Michigan: Blackwell.
- [39] Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing*, 19(1), 3-31.
- [40] Kuchartz, U. (1998). *WinMAX: Scientific text analysis for the social sciences*. Berlin: BBS.
- [41] Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners*. Frankfurt am Main: Peter Lang.
- [42] Matsuda, M. (1999). Interlanguage pragmatics: What can it offer to language teachers? *The CATESOL Journal*, 11(1), 39-59.
- [43] Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- [44] Omar, A. S. (1991). How learners greet in Kiswahili. In L. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning (Vol. 2)* (pp. 59-73). Urbana-Champaign: University of Illinois.
- [45] Ostrom, T. & Gannon, K. (1996). Exemplar generation: Assessing how respondents give meaning to rating scales. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 293-318). San Francisco: Jossey-Bass.
- [46] *Assessing EFL learners' interlanguage pragmatic knowledge* 19
- [47] Roever, C. (2005). *Testing ESL pragmatics: Development and validation of a web-based assessment battery*. Frankfurt am Man: Peter Lang.
- [48] Rose, K. R. (1994). Pragmatic consciousness-raising in an EFL context. *Pragmatics and Language Learning Monograph Series*, 5, 52-63.
- [49] Rose, K. R. (1997). Pragmatics in the classroom: Theoretical concerns and practical possibilities. In L. F. Bouton (Ed.), *Pragmatics and language learning. Monograph Series (Vol. 8)* (pp. 267-295). Urbana-Champaign: University of Illinois.
- [50] Rose, K. R. & Kasper, G. (Eds.). (2001). *Pragmatics in language teaching*. New York: Cambridge University Press.
- [51] Rose, K. R. & Ng, K.-f. C. (2001). Inductive and deductive teaching of compliments and compliment responses. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 145-170). New York: Cambridge University Press.
- [52] Rose, K. R. & Ono, R. (1995). Eliciting speech act data in Japanese: The effect of questionnaire type. *Language Learning*, 45(2), 191-223.
- [53] Sinaiko, H. W. & Brislin, R. W. (1973). Evaluating language translations: Experiments on three assessment methods. *Journal of Applied Psychology*, 57(3), 328-334.
- [54] Takahashi, T. & Beebe, L. M. (1987). The development of pragmatic competence by Japanese learners of English. *JALT Journal*, 8, 131-155.
- [55] Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2), 91-112.
- [56] Yamashita, S. O. (1996). *Six measures of JSL pragmatics*. Honolulu: Second Language Teaching & Curriculum Center of University of Hawaii at Manoa.
- [57] Yoshitake-Strain, S. (1997). *Measuring interlanguage pragmatic competence of Japanese students of English as a foreign language: A multi-test framework evaluation*. Unpublished doctoral dissertation, Columbia Pacific University, Novata, CA.

The Teacher's Performance in Learning Process Management And Chemistry Learning Difficulties Identification

Budi Utami¹, Sulisty Saputro¹, Ashadi¹, Mohammad Masykuri¹, Nonoh Siti Aminah²

¹Chemistry Education

²Physics Education

Teacher Training and Education Faculty
Sebelas Maret University
Jl. Ir. Sutami No. 36A Kentingan Surakarta
Email : budiutami@staff.uns.ac.id

Abstract— This research is aimed at mapping the chemistry learning process activities, and finding out the teachers' difficulties in teaching Chemistry. To achieve the objectives, this research used descriptive method. The data were collected by using questionnaire, interview, observation, and document analysis. The procedure of data analysis was by reducing and verifying the data. The subjects of the research were teacher and students. The results of the research show that: (1) The activities done by the teacher were: a) conducting the introduction, apperception, and motivation which is 88,02%, b) implementing various learning model/strategy which is 89,28%, c) utilizing learning source/various media which is 73, 69%, d) implementing practical activities which is 82,14%, e) guiding group discussion activities which is 91,89%, f) carrying out the evaluation which is 89,19%, (2) The difficulties in learning process experienced by the Chemistry teachers were : a) having limited time allotment to apply the practical method, b) designing the students assignment by themselves, c) creating the media, especially by using computer, d) having the limited ability to optimize the internet in looking for the learning material, e) having the limited time to assess the affective aspect, f) applying the cognitive aspects in analysis stage, synthesis stage, and evaluation stage, g) the students inactivity in discussion, h) having limited time allotment in organizing affective assessment questionnaire, i) lack of tools and materials used in Chemistry Laboratory.

Keywords: *Teacher's Performance, Learning Process Management, Chemistry Learning*

I. INTRODUCTION

In the education field, teacher is one of the important elements. Teachers' role and responsibility definitely decide the success management achievement in education. Teachers are not only assigned to teach the students to accomplish the learning purposes, but also to guide the students, especially to attain education management purposes in every education institution, and to attain the National Education Purposes in general [1].

Teachers have the great role in the successfulness of teaching and learning process. Teaching is not merely about delivering the material, but it is also the purposeful and complex job. To do the duty well based on the skill area, it is necessary to have the sufficient skill level. Being a teacher is not merely about comprehending the material that must be delivered, it also needs to have the ability, competency, and comprehension about knowledge and skill [2].

As stated in the Republic of Indonesia Constitution No. 14 Year 2005 about Teacher and Lecturer, and Government Regulation (PP) No. 19 Year 2005 about National Education Standard that teacher is officially equal with the other professions as the professional worker. As the professional worker, teacher has to fulfill several requirements. Those requirements are the teacher must have academic qualification and 4 competencies: pedagogic competency, professional competency, social competency, and personal competency. Teachers'

competencies are a set of knowledge, skill, and behavior that must be owned, internalized, mastered, and applied by the teachers in doing their professional duty that proved by the work performances.

For the teachers' analysis need as the educators, teachers' ability or teachers' competencies which are mostly related to the effort in increasing the learning process and result can be categorized into four abilities, those are: (a) planning the teaching program, (b) conducting and leading/maintaining teaching and learning process, (c) assessing the teaching and learning progress, and (d) mastering the teaching material in the case of mastering the study field or the subject they taught. Those four abilities are the abilities which must be fully mastered by the professional teachers. Based on the explanations above, teachers' competencies concept can be defined as the basic ability in doing the teaching duties that can be viewed from the ability in planning the teaching and learning process, the ability of doing or managing the teaching and learning process, the ability of assessing teaching and learning process [3].

The objectives of this research are to find out: 1) the teaching and learning activities of Chemistry subject, 2) the difficulties facing by the teacher when teaching Chemistry. The benefits of this research are: 1) to give the information about the teaching and learning process of Chemistry subject in the classroom, 2) to give the information about the difficulties facing by the teachers when teaching Chemistry.

II. RESEARCH METHOD

This research was conducted in Senior High School in Surakarta and the surroundings. The populations of the research were 20 teachers teaching in State and Private School in Surakarta and the first semester of XI grade Science students in the academic year of 2015/2016. The research used descriptive method. Instruments used in this research were questionnaire, observation, and interview. The data collected in this research were analyzed by using data reduction technique, data display, and conclusion.

III. DISCUSSION

The description of teachers' activities mapping in the Chemistry learning process can be viewed in the following Table 1.

Table 1. The description of teachers' activities mapping in the Chemistry learning process

Learning Activities	Average (%)
Conducting introduction, apperception, motivation	88,02
Implementation of various learning model/strategy	89,28
Utilizing learning source/various media	73,69
Implementing practical activities	82,14
Guiding group discussion	91,89
Carrying out the evaluation	89,19

In the Chemistry learning process, the teachers have conducted the apperception and motivation. By using apperception and motivation, teachers' related the Chemistry material with the daily life. Therefore, the students are expected to be motivated in learning Chemistry and to make the Chemistry learning meaningful.

In the learning process, the teachers have implemented the various active learning models. By implementing various active learning models, the students become interested in learning Chemistry. Moreover, they will not feel the boredom, be motivated to pay attention, and be more active in learning Chemistry. The active learning models implemented by the teachers are cooperative learning model, Problem-based Learning model, Problem Solving, Inquiry model, and Contextual Approach.

In the process of learning Chemistry, the teachers have utilized various learning sources, for example, secondary books and internet. However, the teachers remain having the obstacles in utilizing internet and computer in learning Chemistry, for example, limited internet connection and

computer facilities. The teachers may assign the students to utilize the internet facility outside the school.

The learning process will be definitely enjoyable if the teachers are able to package the learning technology. Technology has the important role in deciding the living quality of the people that influences all of life aspects. It also affects the culture quality of the particular country. The teachers of the current century are facing the reality that the students attending the class have had the abundant information they got from the outside school like television and internet [4]. The teachers are required to have the competencies in utilizing the technology especially internet (e-learning). The use of technology in education aims to ease the learning process and to make it accessible for the students so that the learning goals are achieved [5].

The teachers have implemented the practical method in learning Chemistry, but there are the obstacles: limited time causes the teachers rarely conduct the practical activities. Several schools also have the obstacles caused by the limited tools and materials in Chemistry laboratory. The utilizing of the laboratory in the school is very essential to prove the abstract Chemistry substances, so that they will not only be the theories. Furthermore, the utilizing of library and internet in the school as the facilities can enrich the students' and teachers' knowledge to go deeper in the material they are learning [6].

Through the science experiments in the laboratory, the students use the different senses by touching, tasting, moving, listening, smelling, and sometimes testing the material in a controlled way. It help the students to increase the concrete level of thinking to the more complex one [7] that may promote the higher thinking skill in 21st century. Besides observing, communicating is also necessary in the process of building the skill and knowledge. Communication may take the various form of communication including words, action, and symbol graphic to describe a certain condition. It needs the students' help to collect the information through the observation so that it can be shared to others [8]. By having a good communication skill, the students will be able to describe the natural phenomenon in the science class.

Learning result evaluation is one of the most important activities in the learning process. The main purpose of conducting the evaluation (assessment) in the learning process is to gain the accurate information about the level of students' instructional objectives, so that the follow-up steps can be attempted. Assessment can be conducted in the beginning of the learning process, during the learning process and after the learning process. Assessment can be in the form of cognitive, psychomotor, and affective [6]. Based on the result of the interview with the teachers, it can be concluded that the obstacles faced by the teachers are in conducting the affective assessment and cognitive assessment in the level of analysis, synthesis, and evaluation. The best way to decide the critical thinking skill are by using three highest levels of analysis, synthesis, and evaluation [9]. The teachers are not only expected to adopt active method in the learning involving the students directly, but also give the strategy to help the students develop the critical thinking skill and solve the problems in the daily life [10].

IV. CONCLUSION

Based on the result of Chemistry learning process mapping, it can be concluded that: 1) the activities conducted by the teacher are: a) the implementing of introduction, apperception, motivation is 88,02%, b) the implementing of various learning model/strategy is 89,28%, c) the utilizing of various learning source/media is 73,69%, d) the conducting practical activities is 82,14%, 2) the guiding of group discussion activity is 91,89%, f) the conducting of evaluation is 89,19%, 2) The difficulties in learning process experienced by the Chemistry teachers are : a) having limited time allotment to apply the practical method, b) designing the students assignment by themselves, c) creating the media, especially by using computer, d) having the limited ability to optimize the internet in looking for the learning material, e) having the limited time to assess the affective aspect, f) applying the cognitive aspects in analysis stage, synthesis stage, and evaluation stage, g) the students inactivity in discussion, h) having limited time allotment in organizing affective assessment questionnaire, i) lack of tools and materials used in Chemistry Laboratory.

REFERENCES

- [1] Arifin,M. 2005. Strategi Belajar Mengajar Kimia. Malang : Universitas Negeri Malang
- [2] Sanjaya,W. 2006. Strategi Pembelajaran Berorientasi Standar Proses Pendidikan. Jakarta : Kencana Prenada Media

- [3] Balqis,P; Usman,N; Ibrahim,S. 2014. Kompetensi Pedagogik Guru Dalam Meningkatkan Motivasi Belajar Siswa Pada SMPN 3 Ingin Jaya Kabupaten Aceh Besar. Jurnal Administrasi Pendidikan ISSN 2302-0156 Pascasarjana Universitas Syiah Kuala 14 Pages pp. 25- 38
- [4] Amin, M dan Thamrin, A.M. Hubungan Kompetensi Pedagogik Dan Kompetensi Kepribadian Dengan Kinerja Guru. <http://jurnal.untan.ac.id/pdf> diakses tanggal 15 Desember 2015
- [5] Mulyasa, E., (2007), *Korikulum Tingkat Satuan Pendidikan*, Bandung: Remaja Rosdakarya.
- [6] Siregar,R.A. 2012. Analisis Kompetensi Pedagogik Guru Kimia Dan Hubungannya Dengan Hasil Belajar Kimia Siswa SMA Di Kota Padangsidempuan. *digilib.unimed.ac.id*
- [7] Turiman,P., Omar,J., Daud,A.M. & Osman,K. 2012. Fostering the 21st Century Skills through Scientific Literacy and Science Process Skills. *Procedia - Social and Behavioral Sciences* 59 (2012) 110 – 116.
- [8] Bilgin, B. (2006). The effects of hands-on activities incorporating a cooperative learning approach on eight grade students' science process skills and attitudes toward science. *Journal of Baltic Science Education*. 1(9) : 27 – 37.
- [9] Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. New York: Longman.
- [10] Larry P. Nelson, L.P. & Crow,M.L. 2009. Do Active-Learning Strategies Improve Students' Critical Thinking? *Higher Education Studies*; Vol. 4, No. 2; 2014, ISSN 1925-4741 E-ISSN 1925-475X. Published by Canadian Center of Science and Education

COMPONENTS OF SCIENTIFIC ATTITUDE FOR TEACHER OBSERVATION IN PHYSICS LEARNING IN SENIOR HIGH SCHOOL

Elvin Yusliana Ekawati¹

¹Physics Education, FKIP, Sebelas Maret University
elvin.pfisika@gmail.com

Abstract—Components scientific attitude needs to be determined first in the development of assessment instruments scientific attitude by observation. There are various kinds of insights and descriptions related scientific attitude, according to experts in the literature, but not all components can be observed in physics learning in high school. This study aimed to find the content of scientific attitude that can observe in physics learning in high school. The descriptive research method was chosen in order to find the components of scientific attitudes through interviews, observation and questionnaires. The data were drawn from SMA N 1 Surakarta, SMA Al Islam Surakarta, and SMA Batik 2 Surakarta. Quantitative data were analyzed with an Aiken formula, and Qualitative data were analyzed with model interactive from Miles, Huberman & Saldana (2014), for selecting content scientific attitude that can be observed. The results showed that the components of scientific attitude that was evident in physics learning in high school, which are: curiosity, objectivity, suspended judgment, open-mindedness, honesty and hard working.

Keywords: *aiken formul, scientific attitude, teacher observation*

I. INTRODUCTION

Development of scientific attitude is one among the most important outcomes of science education [1]. Science educators have long recognized that scientific attitudes are among the most important outcomes which should result from science teaching [2]. Even it is important, equally as cognitive aspects of science education. Scientific attitude encourages questioning mind and a spirit of inquiry [3]. Therefore, without this, studies of science will only mean acceptance of dogma and will never lead to the development of proper orientation towards various scientific endeavors [4]. Physics is the part of science, and has the same characteristics with science. Scientific attitudes are also important to develop in the learning of physics.

Scientific attitudes defined by Rao, are attributes of an individual who not only behave outwardly in the desirable way towards any scientific endeavor, but also understand why they act as they do so [5]. A scientific attitudes have three basic components: belief, feeling and action. Belief is the cognitive basis of scientific attitudes, which provides a learner several scientific information of scientific phenomenon, eminent scientists, scientific inventions, etc. Central component of scientific attitudes are the feeling towards the belief. Effective science teaching, personality of science teacher, teaching- learning environment etc. plays a major role through which the belief of a learner is converted into his/her feeling towards it. Feeling is associated with emotion, on the basis of which a science learner develops his/her opinion. Third component of scientific attitude is action or behavioral component which is mostly conative. It is the tendency of a science learner to act towards his/her scientific belief in accordance with feeling or opinion. For one reason or another, a people do not or cannot always act the way they feel, but tendency is there [3].

It is now generally acknowledged in the literature that the science related attitudes mentioned in such goal statements do not form a single unidimensional construct. Various writers have attempted classifications [6]. The components of the scientific attitude mentioned in the literature seem to fall into three broad groups. These groups are: (1) general attitude towards ideas and information, such as curiosity, openmindedness, scepticism; (2) attitudes related to the evaluation of ideas and information, generally labelled as critical mindedness and containing such things as objectivity, intellectual honesty and caution in drawing conclusions and making decisions and (3) commitment to particular scientific beliefs such as loyalty to truth, cause and effect relationships, etc. [7].

Judging from their prevalence in the literature, some studies in the measurement of scientific attitudes seem to be most widely accepted by science educators. These studies may be briefly summarized in Tabel 1 below.

Table 1. Summarized of scientific attitudes component

Researcher / Writer	Scientific Attitudes Component
Sood and Sanadhya (1978) adopted by Rao (1996)	rationality, open mindedness, curiosity, aversion to superstition, objectivity of intellectual belief, suspended judgment [5]
Gohit dan Sredevi (2008)	curiosity, rationality, open mindedness, objectivity, suspended, judgment, freedom, superstition, perseverance [8]
Billeh & Zakariades(1975)	rationality, curiosity, open-mindedness, aversion to superstitions, objectivity-intellectual honesty, suspended judgement [9]
Jones & Butts (1983)	concern for evidence; causation; new evidence; sources of information; suspension of judgement; honesty [6]
Pitafi & Farooq (2012)	curiosity, rationality, willingness to suspend judgment, open mindedness, critical mindedness, objectivity, honesty and humility[10]
Yasar & Anagun (2009)	curiosity, skepticism, critical thinking, depending on proofs, persistency, cooperation with others, objectivism and accepting uncertainty [11]
Koslow & Nay (1976)	critical-mindedness, suspended judgement, respect for evidence, honesty, objectivity, willingness to change opinions, open-mindedness, questioning attitude[12]
Harlen (1996)	curiosity, respect for evidence, critical reflection, perseverance, creativity and inventiveness, open mindedness, co-operation with others, willingness to tolerate uncertainty, sensitivity to environment[13]

Although some educators have recognized the scientific attitudes as by-products or concomitant forms of learning there has been a growing tendency to view these attitudes as equal to, or superior to, the knowledge objective of science instruction [14]. Science teachers are becoming aware that if scientific attitudes are to develop from the study of science, they must be taught directly and systematically in the same manner as a mastery of the principles of science is developed [15].

The main purpose of evaluation in the affective domain (scientific attitudes) should be to guide the development and improvement of teaching methods and materials rather than to assign pupil marks [16]. Student scores on valid evaluation instruments can be used as an indication of the extent to which the applied methods and materials encourage the desired behaviors. The preparation of these instruments is facilitated if the objectives are defined in behavioral terms [12]. In other hand, science teachers are not unaware of the need, however, of some valid and reliable test of scientific attitudes [6]. In particular, the knowledge of different operational dimensions of scientific attitude and its measurement might enable a science teacher to identify inclination of science learners towards different endeavours in science. This psychometric consideration of scientific attitude is focus area of the present study. The present purpose of study has also been identified the content of scientific attitude that can observe in physics learning in high school.

II. RESEARCH METHOD

This study is part of a dissertation research that aims to develop an assessment instrument of scientific attitude for teaching physics in high school. The research activities carried out in SMA N 1 Surakarta, SMA Al Islam Surakarta, and SMA Batik 2 Surakarta. Selection of schools with the consideration that the school represents a group of high, medium and low-achieving schools, and the school is conducting scientific activities in learning physics. The data source of this

research are physics teachers, students of class XI and experts who have expertise relevant to this study.

This study is a preliminary study to obtain components and indicators for each component of the scientific attitude which will be developed into an instrument for teacher observation. The research approach used combines qualitative and quantitative approaches. Qualitative data were collected by conducting a literature review a number of documents related to the research results relevant to the scientific attitude. The results of these studies form the basis to observe the scientific attitude of students and interviews with teachers after the observation during the process of learning physics. After that step, the components of scientific attitude are summarized, described, and reviewed by a number of teachers and experts in the field of learning physics.

Summary and description of the scientific attitude components reviewed by delphi technique. This technique is used to collect quantitative data via questionnaires. The scale of questionnaires is a likert scale, with 4 choices of response (very important, important, less important, not important). The reviewer may choose the components of the scientific attitude are important (priority) to be developed in learning physics in high school. After reviewing the summary of the components of a scientific attitude, the next step is involving experts (expert planning of learning physics, expert evaluation of learning physics, expert of fundamental physicist for focus group discussion (FGD). At this stage of discussion to decide the components of scientific attitude is a priority development.

Quantitative data were analyzed using the formula Aiken [17], and qualitative data were analyzed using a model Miles, Huberman, and Saldana, 2014. Data were analyzed interactively through cycles: collecting data, reduction data, displaying data, and verification/drawing the conclusion [18].

III. RESEARCH FINDINGS AND DISCUSSIONS

The study begins with a review of the literature reference that relevant to the scientific attitude. This activity generates a list of components with a description of each component of the scientific attitude. Not all components of scientific attitude on the list above can be observed in the process of learning that takes place in SMA N 1 Surakarta, SMA Al Islam Surakarta, and SMA Batik 2 Surakarta. The results of observation and interviews with teachers of physics suggests that there are several components of the scientific attitude can be observed, namely, curiosity, objectivity, honesty, open-mindedness, rationality, willingness suspended judgment, critical-mindedness, skepticism, critical-mindedness, perseverance, humility, aversion to superstitions. These components are the result summary of components that often occurs during the learning of physics. Some components that are likely to constitute a social attitude not included in the summary list, because it is less representative of the typical characteristics of learning physics based scientific activities.

A. The results of quantitative analysis

For the obtained quantitative data, now filled by 2 expert evaluation of learning physics, 8 expert planning of learning physics, 3 expert fundamental of physics expert . The total number of reviewers is 13 expert. The results of the qualitative data processing using formula Aiken, described in table 2.

Table 2. The results of the qualitative data

NO	Scientific Attitude Component	V	Result
1	Curiosity	0,923	Used
2	Honesty	0,897	Used
3	Objectivity	0,974	Used
4	Open-mindedness	0,846	Used
5	Willingness to suspend judgment	0,744	Used
6	Rationality	0,821	Used
7	Skepticism	0,641	Not Used
8	Critical-mindedness	0,846	Used
9	Humility	0,667	Not Used
10	Perseverance	0,923	Used
11	Aversion to Superstitions	0,641	Not Used

Criteria for deciding component of scientific attitude takes precedence in learning physics, which is the component must have a value V analysis result greater than $V (m = 13) = 0.69$. $V (m = 13)$ is a V from Aiken table for the number reviewer / rater (m) as many as 13 people.

Based on the analysis in Table 2 above, it can be concluded that the components of scientific attitudes that will be developed further, namely curiosity ($V = 0.923$), honesty ($V = 0.897$), Objectivity ($V = 0.974$), open-mindedness ($V = 0.846$), willingness to suspend judgment ($V = 0.821$), critical-mindedness ($V = 0.846$), and perseverance ($V = 0.744$). All of these components has a value V over than 0.69. The components is content that will be the basis for development of assessment instruments of scientific attitude in learning physics in high school. Components that are not used have a V value analysis result is less than $V (m = 13) = 0.69$, namely humility, skepticism, and aversion to superstitions.

B. Results of Qualitative Analysis

The results of a quantitative analysis of the summary description of the components and scientific attitude, is the basis for identifying indicators of scientific attitude. After these steps, the indicators compiled as a material for the process of selecting components of a scientific attitude through discussions in the Forum Group Discussion (FGD). FGD activity was attended by 22 experts representing the expert evaluation of learning physics, a physics lesson planning and basic physics expert. In this discussion, the scientific attitude indicators were selected to decide the components of scientific attitude and revise each scientific attitude indicator.

In the FGD activities, it found similarity component description scientific attitude and indicators for the different components of scientific attitude. Two pairs of components that have similar description and the following indicators: (1) Objectivity and rationality; (2) open-mindedness and critical-mindedness. Experts agree ruled that the components represented by component Objectivity rationality and critical-mindedness component represented by the components of open-mindedness. Overall, the results of focus group concluded that the components are decided to be content in the development of assessment instruments that curiosity scientific attitude, honesty, objectivity, open-mindedness, wich are: willingness to suspend judgment, and perseverance.

In addition to deciding which became contains components in the development of a scientific attitude assessment, there are some results in FGD activities, wich: (1) In general, scientific attitude on learning of physics do not differ with scientific attitude to learning is still in a clump science. Physics and sciences are in a clump sciences (such as chemistry and biology) requires scientific process or scientific activities to learn it. Scientific process at the clump of science have the same procedure, namely to formulate the problem, formulate hypotheses, collect data/ experiments, testing hypotheses and conclude the experimental results. In this process, students gain experience to observe, ask, gather information or conduct an experiment, associate or process information and communicate; (2) Specifically, the difference in the scientific process of learning physics with biology and chemistry lies in: (a) the amount and units used in physics learning more than teaching biology and chemistry; (b) experimental procedures in learning physics in some more complex concepts; (c) measurements must be performed precisely and accurately through multiple data retrieval, because the measurement error will contribute to the analysis results that deviate from the theory; (D) the analysis used in physics is more varied and demanding more complicated calculation.

IV. CONCLUSIONS AND SUGGESTIONS

Scientific attitude is an attitude that needs to be developed in learning at the high school level, especially in learning physics. Students who have good scientific attitude, will be able to solve scientific problems in physics learning. Scientific activities to support the development of a scientific attitude of students, namely: experiment, demonstration, inquiry and report writing. Based on collection and analysis data can be concluded that the scientific attitude components that can be observed in teaching physics in high school, wich are: curiosity, honesty, objectivity, open-mindedness, willingness to suspend judgment, and perseverance. The components of the scientific attitude can be observed and developed by teachers in teaching high school physics. The results in this study may be material to the preparation of assessment instruments and rubrics instruments scientific attitude in terms of the observations by the teacher. Summary results of this study can also be used as a basis for the development instrument of a scientific attitude in the form of self-assessment and peer-assessment.

REFERENCES

- [1] Carin, M. Teaching modern science. Upper Saddle River, NJ: Merrill/ Prentice Hall. In S, 1997.
- [2] Sam S. Blanc, "Review of the General Goals in Science Teaching," Science Education, (February, 1952). 36 :47,
- [3] Mukhopadhyay, R. "Whether aptitude in physics, scientific attitude, and deep approach to study explain achievement in physics significantly – an investigation". International Journal of Humanities and Social Science Invention. 2013, 2 (1), 57-63. www.ijhssi.org
- [4] Lacrose, A.G. & L. "Wisconsin. Test for Measuring Scientific Attitude of 12-th standard students". School Science and Mathematics. 2007
- [5] Rao, D.Bscientific. Attitude vis-à-vis Scientific Aptitude, Discovery Publishing House: N.Delhi.1996.
- [6] Jones, B. & Butts, B. "Development of a set of scales to measure selected scientific attitudes". Research in Science Education, 1983. 13, 133 – 140
- [7] Gauld, C.F. & Hukins, A.A. "Scientific attitudes: A review". Studies in Science Education, 1980, 7, 129-161.
- [8] Gohit, R. K. & Sreedevi, K.V. Scientific Attitude Scale, Constructivism in Science Education, Discovery Publishing House. 2008.
- [9] Billeh, V. Y. & Zakhariades, G. A. "The development and application of a scale for measuring scientific attitudes". Science Education, USA: John Wiley & Sons, Inc, 1975, 59 (2), 155-165..
- [10] Pitafi, A. I. & Farooq, M. "Measurement of scientific attitude of secondary School students in Pakistan". Academic Research International, 2012, 2 (2).
- [11] Yasar, S. & Anagun, S.S., "Reliability and Validity studies of the science teaching course-Scientific Attitude Scale", Journal of Turkish science education, 2009, 4 (2),
- [12] Kozlow, M. J. & Nay, M. A." An approach to measuring scientific attitudes". Science Education, 60 (2), 147-172. USA: John Wiley & Sons, Inc, 1976.
- [13] Harlen, W. The Teaching of Science, second edition. London: David Fulton, 1996.
- [14] Baumel, H. B. & Berger, J.J. "An Attempt To Measure Scientific Attitudes", Science Education, 1965, 49: 267-269,
- [15] Nathan S. Washton, Science Teaching in the Secondary *School*. New York: Harper & Brothers, 1961, p. 39.
- [16] Bloom, B.S., J.T. Hasting, and G.F. Madaus, Handbook on Formative and Summative Evaluation of Student Learning. New York, McGraw-Hill Book Co., 1971.
- [17] Aiken, L. R. "Three coefficients for analyzing the reliability and validity of ratings". Educational and Psychological Measurement, 1985. 45, 131 -142
- [18] Matthew B., A. M. Huberman, J. Saldaña. Qualitative Data Analysis: A Methods Sourcebook and The Coding Manual for Qualitative Researchers. 3rd Edition, SAGE Publications: Los Angeles, London, New Delhi, Singapore, Washington DC, 2014

THE DEVELOPMENT OF PSYCHOMOTOR COMPETENCY ASSESSMENT ON PHYSICS EDUCATION STUDENT OF PALANGKA RAYA UNIVERSITY

Enny Wijayanti

Palangka Raya University

ennywijayanti28@yahoo.com

Abstract—This study aimed to 1) develop psychomotor competency assessment on Physics Education Students' as supplementary psychomotor assessment on Physics teaching and learning, 2) find out construct of psychomotor competency instrument. The development procedure consisted of two stages: development and validation, based of the spiral model from Cennamo & Kalk. The development stage covered the preliminary study, define, design, and demonstrate. The resulted in prototype 1 prescription of psychomotor assessment was validated by focus group discussion and Delphi technique, readability, and expert judgment.. The validation stage covered of developing and delivering, small group tryout, analysis and revision. Content validity expert judgment was estimates through formula content validity coefficient Aikens' V. The construct psychomotor competency was adapted from Trowbridge and Bybee. The construct validity in this tryout was analyzed using software SmartPLS2.0M3 with reliability value psychomotor 0,84. The subjects of this study were students' two semester on Physics Education at University of Palangka Raya. The result of study can be concluded as follows: 1) The instrument of psychomotor competency can be used as supplementary psychomotor assessment on Physics teaching and learning, 2) The result of construct validity of psichomotor competency consist of moving, manipulating and communicating, because of creating unattained.

Keywords: *psychomotor competency assessment*

I. INTRODUCTION

Teaching is a process of helping an individual to build his or her own knowledge. In other word, teaching is not about transferring knowledge from a lecturer to students, rather it helps the students to be able to construct their own knowledge through their activities in studying the phenomenon and objects that suit their interests. Facilities and infrastructure improvement will be required to support the activities of the students as well as the environment that may stimulate the students to engage in critical dialogue.

In this process, the responsibility of the lecturer is merely as an academic guide who actively asks the students, stimulates the students' thoughts, creates the issues, lets the students express ideas and concept the thoughts and the ideas of the students, while explaining whether the thought is right or wrong.

. The material on Physics learning consists of two components, namely: the scientific component form and the process scientific component form. This second component give in consequences in meaning nor application on learning activity. The scientific form studied use logic structure and systematic, the shape of the scientific form, namely: concept, principle, theory, and law. Whereas the scientific process studied through empiric observation and logic reasoning.

This research aim to: (1) develop psychomotor competency assessment as a supplementary assessment on Physics learning existing, (2) find out construct of instrument psychomotor competency assessment. Assessed psychomotor competency inthe research development is related to the activities in the laboratory that were conducted by the students. Laboratory activities are the application of the theories that the students have learned. Since the laboratory activities in physis learning have a crucial role, thus, a lecturer or assistant should be able to

plan and manage these activity properly. Without good planning and management, these activities would not be going to work well to support the goal of leaning objectives as expected.

Physics laboratory activities aim to: (a) developing skill in observation, recording of measurement data and manipulating tool that is required was well as creating simple tool; (b) working carefully, analyzing the data accurately, correctly, and objectively; (c) developing the power of logical thinking critically and accurately; (d) developing a scientific attitudes. This does accor to Chiappetta & Kobala 2010, p. 213., Collete & Chiappetta, 1994, p. 198.

Acording Carin & Sund (Tim Pustaka, 2007, p. 283), defines science as asystematic knowledge and arranged regularly, universal, and as a form of data collection from the result of observation and experimentation. According to their opinion, it can be concluded that science covers four main elements, such as: (1) the attitude: curiosity about natural phenomena, objects, living beings as well as the causal relationship that may raise new problems which can be solved through correct procedures, (2) process: troubleshooting procedures through the scientific method, (3) product: facts, principles, theories and laws, (4) application: the application of scientific methods and concepts of science in everyday live.

Therefore, in learning physics, those four elements mentioned above should be available so that the attitudes of students and skills grow well along with the cognitive competence achived. Learning physics is expected to become a vehicle for students to learn about nature as well as the prospect of further development in everyday life. Learning physics emphasizes direct experience to develop competence so that the students able to understand the natural surroundings through the process of "finding out" and "doing". The skill in finding out or doing is called the inquiry process skill or "inquiry skill". By the careful process of skills, attitudes, and values that developed include: curiosity, honest, patient, open, critical, careful, diligent, disciplined, caring for the environment.

Based on the description above, it can be concluded that the learning process of physics can not be separated from the laboratory avtivities. The experience gained in the laboratory activities is not only based on *mind-on* but also *hand-on*. Thus, it will help the students to understand the concepts, principles, theories and laws of physicss in particular. In addition, the students directly practice using the scientific method through the laboratory activities. Therefore, the approach to laboratory activities constitutes a suitable approach used in the study of science in general and physics in particular (Collete & Chiappetta, 1994, p. 198). This is in accordance with the opinion of Trowbridge & Bybee (1986, p. 39), who stated that science is both a body of knowledge by a process.

Psychomotor competency assessment used in the practical activities of the Physics Education Program in Palangka Raya University is based on the results of the practical work, while practical work observation assessment is using *check list*. Psychomotor competency uses a *check list*. The preparation and its use is simple. However, it has a weakness since it uses dichotomous variable (yes or no). *Check list* instrument cannot be used to identify students abilities in detail or unreliable to detect the level of understanding and skill of the students. Consequently, an assessment that can give more objective result is required as well as more precisely in measuring the psychomotor.

The developed complementary assessment was adopted from Trowbridge & Bybee (1986, p. 135) theory of measurement of psychomotor domain. Psychomotor competency assessment observes the laboratory activities using the *rating scale*. *Rubric* is used the reference in providing the score of the students' performance while doing practical work. The preparation of the items of instruments by outlining the psychomotor domain components. These components are described into skill indicators that will be measured. Furthermore, the indicators are formulated into some items or aspects that are assessed. The use of an instrument to measure the mastery of developed psychomotor competency will get more objective and more fair measurement results.

II. THEORY REVIEW

Literally, competence is the ability to master something (Balai Pustaka, 1994, p. 516). According to Trowbridge & Bybee (1986, p. 131-135), competence is the ability of performance on the cognitive domain, affective and psychomotor learning results, as an observable learning result (Gronlund & linn, 1990, p. 11), shaped behavior change (Reigeluth & Stein, 1983, p. 9). According to the statements above, psychomotor competence is the ability of the psychomotor domain performance, as a result of learning can be observed through the skill and behaviors that successfully demonstrated.

According to Trowbridge & Bybee (1986, p. 135), psychomotor competence is related to the ability in motion, psychomotor competency component, including the following: *moving*, *manipulating*, *communicating*, and *creating*. Trowbridge & Bybee argued that the peculiarity in the subjects of science, including physics, in psychomotor domain related to the results which involving ways to manipulate tools (instruments). The component measured in the psychomotor competence in science is different from the psychomotor on subjects of "Physical Education". Therefore, it has to be measured in accordance with the field of science through the laboratory activities at the time doing practical work. The material used in this study is Simple Harmonic Motion on the spring and simple pendulum. Psychomotor component and description of each component is presented in Table 1.

Table 1. Description of Psychomotor Components

No	Psychomotor Components	
	Psychomotor Component	Description
1	<i>Moving</i>	Referring to a number of body movements that involve the coordination of physical movements. The operational verb that used to formulate indicators of achievement are: carry, organize, follow, selecting, placing.
2	<i>Manipulating</i>	Referring to the activity which includes coordinated patterns of some movements which involving parts of body. The operational verb that used to formulate indicators, are: stringing, transform, connect, calibrate, mixing, weighing, operate, heating, repair, adjust, build.
3	<i>Communicating</i>	Referring to the meaning of activity that presents ideas and feelings to be known by others. The operational verb which can be used to formulate indicators of achievement should include: analyze, describe, explain, create graphics, create tables, conclude, shows, discuss, compose, record.
4	<i>Creating</i>	Referring to the process and the resulting of work performance of new ideas. Creations in science subjects usually need some combination of movement, manipulation and communication in generating new results that are unique. The operational verbs that can be used, are: create, design, plan, synthesize, construct, generalize.

Rating categories used in the development of psychomotor competency instrument, namely: Very Good (VG) = 5; Good (G) = 4; Enough (E) = 3; Less (L) = 2; and Very Less (VL) = 1. Based on components, indicators, aspects assessed, so then will be specifying of psychomotor competency instrument, which are presented in Table 2.

Table 2. Specifying of Psychomotor Competency Instrument

No	Specifying of Psychomotor Competency		
	Psychomotor Components	Term for Objectives	Behavioral Objectives
1	<i>Moving</i>	1. Following the determined working procedures 2. Choose tools and materials 3. Clean and replace science material.	1. To follow working procedure 2. To choose tools and materials accuracy. 3. To clean and replace material
2	<i>Manipulating</i>	4. Set the tools 5. Operate experiment 6. Do the observation	4. To set up laboratory equipment quickly 5. To operate instruments quickly 6. To do the observation And experimentation precision
3	<i>Communicating</i>	7. Write the result 8. Check the analysis 9. Make a chart	7. To accurate in writing the result 8. To accurate in analyzing the data 9. To graph data accurately

		10. Conclude the result 11. Report data 12. Listen to other students 13. Ask questions about problems	10.To conclude the result precision 11.To report data accurately 12.To listen to other students 13.To ask question about problems
4	<i>Creating</i>	14.Create different ways of solving problems 15. Plan ways to solve problems	14.To create different ways of solving problems 15.To plan ways to solve problems

Table 2 above shows that the number of items of psychomotor competencies questions which developed are fifteen, and the formula is still general. The items will be measured than will formulated specifically, adjusted to the materials to be used.

III. RESEARCH METHODOLOGY

This study is an experimental development which accompanied by a quasi-experimental. The development of psychomotor competency assessment integrates the learning outcomes measurement with the overall learning process. The development procedure consisted of two stages: development stage and validation stage, based of the spiral model from Cennamo & Kalk (2015, p. 6), then known as the five phases of development, namely: (1) define, (2) design, (3) demonstrate, (4) develop, and (5) deliver.

The development stage, data are collected by interviews, observations, questionnaires, documentation, that was validated through focus group discussion and Delphi technique. The development stage describes about methodological aspects of conceptual development, but previously preliminary study is conducted to know the instruments and the assessment used in practical activities, and study of relevan theory. Furthermore, formulating, designing, grating, observation instrument of practical training, rubrics, and the assessment system. This activity results the prescriptions of psychomotor competency assessment, which is referred as a prototype 1.

The validation stage covered of developing, and delivering. The prototype one is tested the readability, review experts, the try out of small group which is not conducted, the tests is conducted as an instrument validation test empirically. The empirical try out is the product testing activities. Activity in the second stage include analysis and revision, test of content validity coefficient Aikens' V, which produced prototype 2. The testing of psychomotor competency instrument validation using Quasi-Experimental design is adopted from Creswell (2005, p. 314). Subject try out are students of second semester who are taking Basic Physics II.

A. Data Analysis Technique

Analysis of pre-try out to estimate whether all items psychomotor instruments have been valid by content. The testing of content validity use formula content validity coefficient Aikens' V (1985, p. 132 – 134).

$$V = \frac{\sum s}{n(c-1)} \quad (1)$$

Explanation:

s = r – lo (score of rater – lowest score)

r = assessment score of rater

c = number of category

n = number of rater

lo= score rating minimal

The criteria used to estimate these items valid by content, is by comparing values of V count must be greater than the value of V Table Right- Tail Probability for Selected values of the Validity Coefficient (V), at

$\alpha = 0.05$, category of rating four, and the number of raters are six expert. To calculate the amount of inter-rater reliability coefficient was adopted on Aikens' (1980, p. 959), as follows.

$$R = 1 - \frac{\sum_{j=0}^2 \sum_{i=0}^2 n_{ij} |i-j|}{N(c-1)} \quad (2)$$

Explanations:

n_{ij} = number of cels in matrix 3x3

N = number of rater

C = number of category

The try out analysis is analyzed using SmartPLS 2.0 M3 program. Reliability test is performed to test the construct, to prove the accuracy of the consistency and instrument accuration in measuring the construct. Formula for calculating the composite reliability uses the formula developed by Werts., Linn., & Joreskop (Ghozali, 2012. p. 80) as follows.

$$\rho_c = \frac{(\sum \lambda_i)^2 \text{var } F}{(\sum \lambda_i)^2 \text{var } F + \sum \theta_{ii}} \quad (3)$$

Where: λ_i : factor loading; F : factor variance; θ_{ii} : error variance = $1 - \lambda_i^2$

The formula of psychomotor competency mastery level is adopted from Mardapi (2012, p. 107), the criteria used which are presented in Table 3.

$$\text{Mastery Level} = \frac{\text{Total score obtained}}{\text{Maximum Score}} \times 100\% \quad (4)$$

Table 3. Category of Psychomotor Competency Mastery

No	Category of Psychomotor Competency Mastery	
	Level of Mastery	Category
1.	90% - 100%	Very Good
2.	80% - 89%	Good
3..	70% - 79%	Enough
4.	< 70%	Less

IV. DISCUSSION AND RESEARCH FINDING

Results of pre-try out psychomotor competency instruments are validated by six expert. The calculation results of content validity expert judgment are presented in Table 4.

Table 4. Results of Pre-try out Stage by Content Validity Coefficient

No	Table Content Validity Coefficient		
	Validated Aspects	Value V count	Value V table
1.	Assesment observation format of laboratorium activity.	0.89	0.78
2.	Compatibility between each component and aspect that is assessed overall	0.96	0.78
3.	Compatibility of moving components between each items in that aspect.	0.93	0.78
4.	Compatibility of manipulating components with items in that aspects.	0.93	0.78
5.	Compatibility of communicating components with items in that aspects.	0.96	0.78
6.	Compatibility of creating aspects with items in create aspects.	0.93	0.74

Based on Table 4 above, it can be concluded that all the items of psychomotor competence instruments, value $V_{\text{count}} > V_{\text{table}}$. In conclusion psychomotor competency instruments are valid by content. Furthermore, rubric for observation assessment psychomotor competencies have to be validated by content by validator, the results show that all the items produce $V_{\text{count}} > V_{\text{table}}$. These results suggest that the rubric observation assessment laboratory activities is valid by

content. Reliability of inter-rater in try out phase is calculating using (2) above, and the calculation results are presented in Table 5.

Table 5. Reliability of Inter-Rater in Pre-Try Out

No.	Reliability Inter-Rater Table		
	Instrumet	Coefficient Score of Inter-Rater	Explanation
1.	Psychomotor competencies	0.94	Reliable
2.	Assessment Observation Rubric	0.90	Reliable

After all instruments are valid by content, than performed try out for the instrument or validation test. The subject try out is a second semester students who are being programmed Basic Physics II course, as many as 51 peoples. The try out results are analyzed using software SmartPLS 2.0M3 to see the construct validity and reliability, with composite reliability value psychomotor 0,814. The statistic used are weighted regression (weight) which basically is a standardized regression coefficient (path coefficient). Tested using the t-test, as the criteria is used critical value of t-test at significance level (α) = 5%, amounting to 1.96. Result for Outer Weight psychomotor competence with a value of T-Statistic less than 1.96 is stated disqualified. T-Statistics value in *creating* components is less than 1.96. The results of the psychomotor competencies try out after tested are presented in Table 6.

Table 6. Distribution Items of Psychomotor Instruments

Psychomotor Components	Items of questions	
	First Instrument	Last Instruments
Moving (P1)	1, 2, 3	1, 2, 3
Manipulating (P2)	4, 5, 6	4, 5, 6
Communicating (P3)	7, 8, 9, 10, 11, 12, 13	7, 8, 9, 10, 11, 12, 13
Creating (P4)	14, 15	-

Table 6 above shows that the items of *creating* components after try out (validation test) are fall, because it do not qualify. The results psychomotor competencies mastery are presented in Table 7.

Table 7. Psychomotor Competency Mastery

No.	Psychomotor Competency	
	Psychomotor Components	Rating Mastery (%)
	Moving (P1)	82
2.	Manipulating (P2)	78
3.	Communicating (P3)	84

Results of rating mastery in *moving* component is 82%, indicating good, while the result of *manipulating* component is 78%, sufficient category, and the results of *communicating* component is 84%, good category.

Based on the results of psychomotor competence validation test, shows that there is a problems of student in Physics Education Program, Palangka Raya University. Which is creations ability of generally of inter related physics concepts, formulate new questions based on experimental results and applying the techniques of experiments on new problems when do the practical activity. This shows that the purpose of laboratory activities as expected has not been achieved. Therefore, this information is useful for all parties, involved, institution, and all lecturers in the Physics Educational Program of Palangka Raya University.

The physics learning is less in creating the learning atmosphere which stimulates students to be creative, facilities and infrastructure in the laboratory is limited in quantify and quality. Therefore, all related parties should work together to plan and find the best solution so in the future through the laboratory activities appear the better students creativity. Student involvement in this issue show that most of the students are spending less time to learn well, this is a positive feedback fo students.

V. CONCLUSION

The conclusion which resulted on psychomotor competency assessment development research, are: (1) Psychomotor competencies assessment is produced can be used as a complementary assessment of existing psychomotor competencies; (2) The results of construct validity analysis, indicates that students mastery of *creating* components still lacking. *Creating* components associate with create, relates with formula the generalization of model, formulate new problems based on experiment. The study of this development, still in validation test stage, yet a final product. Further analysis and revisions should be made to the validation results to correct the deficiencies.

REFERENCES

- [1] Aiken, L. R. (1980). Content validity and reliability of single item or questionnaires. *Educational and Psychological Measurement*, 40: 955-959.
- [2] Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45: 131-142.
- [3] Cennamo, K., & Kalk, D. (2005). *Real world instructions design*. Canada: Thomson Wadsworth.
- [4] Chiappetta, E. L., & Kobala, T. R. (2010) *Science instruction in the middle and secondary school*. New York: Pearson Education, Inc.
- [5] Collete, A. T., & Chiappetta, E. L. (1994). *Science instruction in the middle and secondary school* (3rd ed.). New York: Macmillan Publishing Company.
- [6] Creswell, J. W. (2005). *Educational research planning, conducting, and evaluating quantitative and qualitative research*. New Jersey: Pearson Education, Inc.
- [7] Ghazali, I., & Latan, H. (2012). *Partial least squares konsep, teknik dan aplikasi menggunakan SmartPLS 2.0 M3*. Semarang: Badan Penerbit Undip.
- [8] Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th.ed). New York: Publishing company.
- [9] Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- [10] Reigeluth, C. R., & Stein, S. F. (1983). *The elaboration theory of construction*. New Jersey: Publisher, Hillsdale.
- [11] Reigeluth, C. R., & Stein, S. F. (1983). *The elaboration theory of construction*. New Jersey: Publisher, Hillsdale.
- [12] Trowbridge, I. R., & Bybee, R. W. (1986). *Becoming a secondary school science teacher* (4th.ed). Columbus: Charles E Merrill Publishing Company.

Implementation of Authentic Assessment in Bahasa Indonesia Subject for Senior High School in West Sumbawa

Eny Rusmaini¹

¹Magister Penelitian dan Evaluasi Pendidikan, Program Pascasarjana
Universitas Negeri Yogyakarta
Enyrusmaini16@gmail.com

Abstract— This study aims to describe (1) teachers' understanding about the authentic assessment procedures in the learning of *Bahasa Indonesia*, (2) The form of authentic assessment being applied, (3) the constraints that are experienced by teachers in implementing authentic assessment, and (4) the efforts to overcome the constraints that are experienced by teachers in implementing authentic assessment. This study is evaluation research using mixed method approach. Discrepancy model is used as evaluation model in this study. The subject of this study is senior high school located in *Sumbawa Barat* which implementing authentic assessment. The objects of this study are students and *Bahasa Indonesia* teachers. The data are obtained by questionnaire, interview, document analysis, and observation. The data from questionnaire are analyzed quantitatively, while the data from interview, document analysis, and observation are analyzed qualitatively. The results of this study show that all of *Bahasa Indonesia* teachers have applied the authentic assessment in learning and teaching process. Teachers apply the authentic assessment in vary model, which tends to creativity factor. Teachers also have the good effort in order to take steps properly in implementing authentic assessment. Nonetheless, teachers face so many kinds of constrains in implementing authentic assessment. The teachers' efforts adjusted to the constrains faced by them. This indicates that teachers trying to implement authentic assessment properly and suitably to the applicable curriculum.

Keywords: *authentic assessment, bahasa Indonesia subject*

I. INTRODUCTION

According to regulation of ministry of education and culture No.66 in 2013, authentic assessment is the assessment which is conducted comprehensively to evaluate the input, process, and output in learning. Authentic assessment is more considered in assessing both the process and the result. Therefore, the students' performance in learning process can be assessed objectively, honestly, and process-orientated. (Burhan Nurgiyantoro, 2015)

According to Burhan Nurgiyantoro (2015), authentic assessment focus on the leaning ability to demonstrate the real students' knowledges. The assessment activity is not merely inquiring or intercepting what have been gained in learning process, moreover, the real performance which has been mastered by students. Mastering the performances, students are required to apply the knowledge in real life, in the school or out the school.

Implementation of authentic assessment has been implemented since The School-Based Curriculum (KTSP). However, in 2013 Curriculum, the authentic assessment is more emphasized. The implementation of authentic assessment in *Bahasa Indonesia* needs more time for teachers to fill the format, i.e. cognitive, affective, and psychomotor, also recapitulation of student scores at the end of learning. To ease the assessment, teachers need to memorize the whole students' name in order to give the proper value to them. The information of student's development that is gained from the listening skill, speaking skill, reading, skill, and writing skill, can be assessed by teachers if they already have known the procedures in authentic assessment. This is because the results of authentic assessment are able to determine what treatment should be given to the student teacher. Teachers can carry out further improvement of the results of student competency achievement.

II. METHOD

This study is evaluation study by using the discrepancy model of evaluation which was developed by Provus (Fitzpatrick, 2011: 155). This study was held in the Senior High School which implement the 2013 Curriculum in Sumbawa Barat. This study is limited to teachers' understanding of the authentic assessment, implementation of authentic assessment, and authentic forms of assessment in *Bahasa Indonesia* to tenth and eleventh grade senior high school students. This study is mixed method, which questionnaire is used to obtain the quantitative data and observation, interview, and document analysis are used to obtain the qualitative data.

Documentation technique is conducted to analyze teachers-made lesson plan in *Bahasa Indonesia*. The questionnaire is intended to determine the students' perception and application of authentic assessment. The sample of this study is 60 students of tenth and eleventh grade students of *SMA Negeri 1 Jereweh* and 60 students of tenth and eleventh grade senior high school students of *SMA Al-Ikhlas*. The scoring of the questionnaire will be calculated its the tendency level using the criterion as follow.

Table 1. Kriteria Evaluasi

No	Skor	Category
1	$X \geq \bar{X} + 1SB_x$	Highly Suitable
2	$\bar{X} + 1SB_x > X \geq \bar{X}$	Suitable
3	$\bar{X} > X \geq \bar{X} - 1SB_x$	Unsuitable
4	$X < \bar{X} - 1SB_x$	Highly Unsuitable

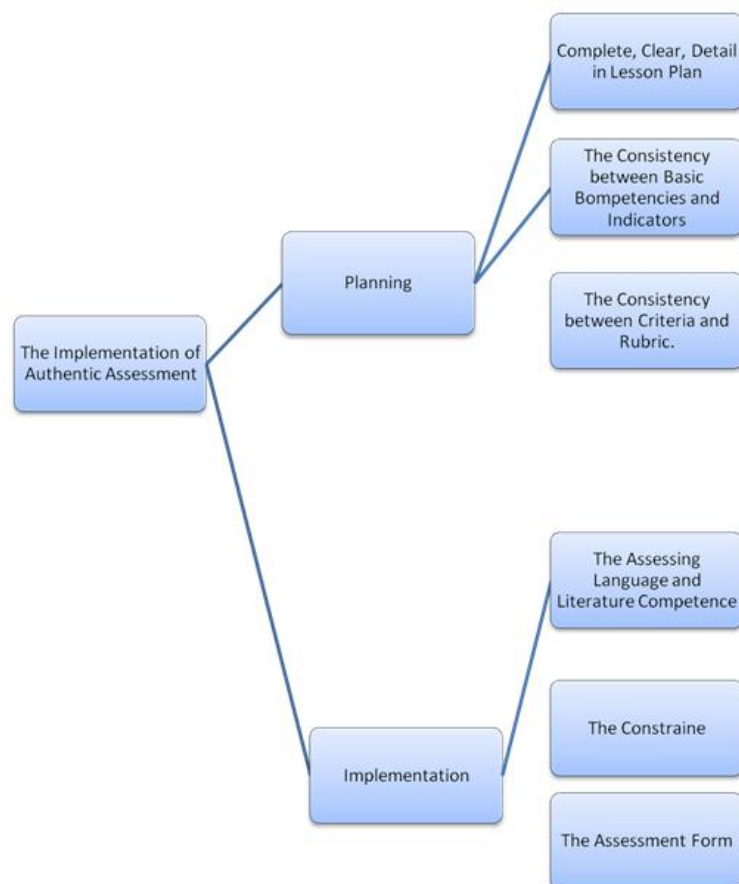


Figure 1. Variable, Dimension, and Indicator of the research

Data collection through interviews were conducted to obtain the additional information as well as data that can triangulate the two previous techniques. The interviews were conducted to three *Bahasa Indonesia* teachers on how the assessment held, the constraints faced, and the efforts to overcome the obstacles made by the teacher. The data that are collected through interviews, can be reduced. The data which are relevant to the application of authentic assessment are used as supporting data.

III. RESULTS AND DISCUSSION

According to the research questions, the research includes teachers' understanding on the implementation of authentic assessment, whether authentic assessment has been implemented or not, the form of authentic assessment that is often implemented, the constraints of applying authentic assessment, and the efforts to overcome the constraints of applying authentic assessment. The results of the study described as follows;

A. The Teacher's Understanding

The teachers' understanding of authentic assessment implementation is known from the document analysis, i.e. syllabus and lesson plan used by teachers in teaching *Bahasa Indonesia*. The result of document analysis shows that the whole analysis subject has been implemented the authentic assessment and attaches in their syllabus and lesson plan.

B. The Implementation of Authentic Assessment

The result of questionnaire analysis shows that the teachers have been implemented the authentic assessment. This is also supported by the interview research which stated that the all teachers have been implemented the authentic assessment because it is demanded in 2013 Curriculum. However, the discrepancy score between the teachers' questionnaire and students' questionnaire is 1.91, related to several descriptors that cannot be measured. This is obtained from the score of questionnaire and maximum score, i.e. 2.089 out of 4.

The planning indicators of authentic assessment is used to gather the information about the completeness and clarity of lesson plan, consistency between basic consistency with indicators, and consistency between the criteria of the scoring rubric used by teachers. This indicator reaches a score of 3.333 out of 4 with the discrepancy score of 0.667. It means that the lesson plan made by teachers can be classified as complete, detail, and clear.

As with the previous indicator, the second indicator is the implementation of authentic assessment that has a total score of 3.333 out of 4 with a value gap of 0.667. This means that in terms of the implementation of authentic assessment on *Bahasa Indonesia* has been suitable to measure students' language and literature competence in terms of both process and results.

According to the result of observation, the all observed teachers has been implemented the authentic assessment. For example, in the lesson material of tenth grade anecdote text, teachers assess in writing competence. Students are asked to rewrite the anecdote from the example given by teachers. The learning in *Bahasa Indonesia* has been improved since the implementation of 2013 Curriculum, including the material which is more likely based on reading text. The materials for tenth grade students in second semester are divided classified in three categories; negotiation text, anecdote text, and explanation text. The core activities in learning *Bahasa Indonesia* for tenth and eleventh grade students are more focus in observing, asking, exploring, associating, and communicating.

According to core activities above, the first is observing where students need to read the anecdote text. The second is asking; students ask the structure of the anecdote text, the arrange the question according to the anecdote text. The third is exploring, where students find out the main structure and the rules of anecdote text (abstract, orientation, crisis, response, and coda). The next is discussion and interpretation among students about the structure of anecdote text. The last is communicating by doing presentation and assessing the validity and accuracy from the conclusion of each team dealing with anecdote text.

The assessment in process is less attention because the teachers are more focus to students' assignment. This can be concluded from the observation through teachers that the teachers are less understanding the concept of authentic assessment in *Bahasa Indonesia*. The concept of authentic assessment is more focus in the process than result. According to the result of interview, teachers state that it needs more time in assessing the process of students.

Besides the teachers do not understand the concept of authentic assessment, they also said that they have been implemented the authentic assessment but felt that its application is not maximum because there are constraints faced. There are also teachers who implement recently. This was confirmed through analysis of teacher's syllabus and lesson plans.

C. The Form of Authentic Assessment

The forms of authentic assessment are varied, since the model of assessment is appropriate to the essence of authentic assessment. The forms of authentic assessment which are implemented by *Bahasa Indonesia* teachers in Sumbawa Barat, i.e. performance assessment, peer assessment, open question, writing sample text, written assessment, and portfolio assessment.

There are six kinds of assessments used by teacher, meanwhile the form of authentic assessment is varied. This shows that the teacher's understanding in authentic assessment is still not enough, so the teachers must be more creative in implementing the form of authentic assessment. According to the result of questionnaire analysis to determine the form of authentic assessment used by teachers, results the discrepancy score 1.18 which obtained from the difference score between maximum score, 4, and questionnaire score, 2.82.

D. Constraints

Constraints are obstacles which are experienced in implementing authentic assessment. The extinction of obstacles would cause the application of authentic assessment less than the maximum. Constraints which are experienced by teacher of *Bahasa Indonesia* teachers in senior high school in Sumbawa Barat, according to interviews, include time constraints, limited facilities and infrastructure, the difficulty of implementing authentic assessment on certain competencies, the difficulty of implementing certain forms of authentic assessment, and constraints of learners.

IV. CONCLUSION

According to the result of analysis and discussion above, it can be concluded that the implementation of authentic assessment in *Bahasa Indonesia* for senior high school in Sumbawa Barat are classified as suitable by reaching the total score of 2.089 out of 4, and the discrepancy is 1.91. Based on this conclusion, the recommendations can be taken into consideration for the parties engaged in the world of education are:

- a. The need for further research regarding the effect of the implementation of authentic assessment with the motivation and learning outcomes of students.
- b. Teachers should always learn to improve understanding and ability in conducting an assessment in accordance with the demands of the curriculum in force.
- c. Teachers should always practice, creative, receptive to correction, and is active in such MGMPs teacher conferences, so that difficulties can be minimized teacher assessment.

REFERENCES

- [1] Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R., (2011). *Program Evaluation Alternative Approaches and Practical Guidelines* (4th ed.). New Jersey: Pearson.
- [2] Gulikers, Judith T.M, Bastiens, Theo J, Kirschner, Paul A. (2004) A Five-Dimensional Framework for Authentic Assessment. *Journal of Educational Technology, Research and Development*, 52, 67-86.
- [3] Kunandar. (2014). *Penilaian Otentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013)* (Ed.Rev). Jakarta: Rajawali Press.
- [4] Lund, Jacalyn. (1997). Authentic Assessment: It's Development and Applications. *Journal of Physical Education, Recreation & Dance*. 68, 25-40.
- [5] Mardapi, D (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendikia Press.
- [6] Nurgiyantoro, Burhan. 2011. *Penilaian Otentik dalam Pembelajaran Bahasa*. Yogyakarta: Gajah Mada University Press.
- [7] Reynolds, C.R, Liwington, R.B, Willson, V., (2009). *Measurement and Assessment in Education* (2nded.). New Jersey: Pearson.

Summative Assessment Design through the PjBL to Improve Students' Higher-Order Thinking Skills

Erlida Amnie

Master of Physics Education, Lampung University
early_da@yahoo.co.id

Abstract—This study is a preliminary stage in developing summative assessment instrument through project-based learning (PjBL) to improve students' higher-order thinking skills (HOTS). The purposes of this preliminary study are (1) to determine the level of the needs of teachers and students as well as the developments in the use of summative assessment of learning that takes place in schools; (2) to design the summative assessment through the PjBL in improving students' higher-order thinking skills. The samples are taken through random sampling. The results of this study were obtained from literature review and survey questionnaires to five Physics teachers and forty high school students in Bandar Lampung in the even semester of the 2015/2016 academic year. Based on these survey results, it was found that teachers often asses the students' through summative assessment, but the teachers do not use assessment through project-based learning in measuring students' higher-order thinking skills. This fact indicates that the development of summative assessment in improving students' higher-order thinking skills is required. Summative assessment to improve higher-order thinking skills conducted at each phase of PjBL. The phases to perform are the project determination, project design, project implementation, project process and results evaluation, preparation of reports and presentations, and project completion through monitoring by teachers.

Keywords: *summative, assessment, PjBL, HOTS*

I. INTRODUCTION

Assessment is a part of learning process which is continuous though daily activities. Assessment is a critical piece of differentiated instruction as it helps to identify the most effective strategies and activities that will encourage student learning. The result of an assessment is the initial process before making a decision on the next steps in improving the ability to understand student learning.

Assessment is the process of gathering data. More specifically, assessment is the ways instructors gather data about their teaching and their students' learning [1]. The data by different forms of assessment such as: pre-tests, observations, and examinations. The next step after gathering the data is evaluation. It draws on one's judgment to determine the overall value of an outcome based on the assessment data. The last step is the decision making process by designing ways to improve the recognized weaknesses, gaps, or deficiencies.

Assessment has more than one purpose and many interested parties. Assessment provides information that can help improve students' learning and help teachers in teaching [2]. Assessment is an integral part of learning. It is easy to believe that good assessment takes into account learning styles, strengths, and needs. It clearly articulated learning destinations of the students.

Doing this well requires new assessment development models that incorporate close collaboration between curriculum designers and assessment developers to ensure tight alignment and seamless integration of assessment and instruction [3]. NSW syllabuses and support materials promote an integrated approach to teaching, learning and assessment. *Assessment for learning*, *assessment as learning* and *assessment of learning* are approaches that can be used individually or together, formally or informally, to gather evidence about student achievement and to improve student learning [4]. Assessment for, of and as learning consider how to intend on the variety of tasks and assignments that are used in the classroom.

Assessment for learning, known as formative and diagnostic assessments. Assessment for learning is the use of a task or an activity for the purpose of determining student progress during a unit or block of instruction. Assessment of learning is the use of a task or an activity to measure, record and report on a student's level of achievement in regards to specific learning expectations. These are often known as summative assessments. Assessment as learning is the use of a task or an activity to allow students the opportunity to use assessment to further their own learning [5].

Summative assessment is an attempt to summarize learning activities undertaken by a student at a certain time [6]. If assessment is a single process with more than one function, then the process cannot logically be divided and the process is the same regardless of the function [7]. Summative evaluation is aimed at assessing if the person or program has fulfilled the stated goals [8]. Summative assessment centers on summing up or summarizing achievement of students, classes, schools, and so forth [9]. Summative assessment methods enable giving grades reflecting students' performance [10]. The way to realize students' achievement and performance is rubric. That rubric is used for finishing a project that will be found through project based learning (PjBL).

The core idea of project based learning is that real-world problems capture students' interest and provoke serious thinking as the students acquire and apply new knowledge in a problem-solving context [11]. In the project-based learning, by solving different problems it is possible to develop creative ideas while improving highly developed skills [12].

The learning process is instrumental in improving students' comprehension. Among others is the students' higher-order thinking skills (HOTS). As described by [13], students are required not only to have a lower-order thinking skills but also the higher-order thinking skills (HOTS), so the students have to get used to facing problems that require higher-order thinking skills (*siswa dituntut tidak hanya memiliki keterampilan berpikir tingkat rendah atau lower order thinking skills, tetapi sampai pada keterampilan berpikir tingkat tinggi atau higher order thinking skills (HOTS). Sehingga siswa harus terbiasa menghadapi permasalahan yang memerlukan higher order thinking skills.*)

Higher-order thinking skills (HOTS) is a process that is not merely to memorize and recall the known information. Higher-order thinking skills is the ability to connect, manipulate, and transform knowledge and experience possessed to think critically and creatively in order to determine decisions and solve problems in new situations (*Kemampuan berpikir tingkat tinggi (Higher Order Thinking Skill – HOTS) merupakan proses berpikir yang tidak sekedar menghafal dan menyampaikan kembali informasi yang diketahui. Kemampuan berpikir tingkat tinggi merupakan kemampuan menghubungkan, memanipulasi, dan mentransformasi pengetahuan serta pengalaman yang sudah dimiliki untuk berpikir secara kritis dan kreatif dalam upaya menentukan keputusan dan memecahkan masalah pada situasi baru*) [14]. HOTS become essential as it can assist them to complete their assignments and learn the subject [15].

Summative assessment in measuring higher-order thinking skills is still rarely used. The assessment requires the preparation of technical instruments that may lead students to think critically and creatively. Based on the explanation, the following preliminary research is done by focusing on the analysis of summative assessments need to increase higher-order thinking skills of students in learning. Before knowing what kind of assessment is needed, a preliminary research should be carried out related to analysis of needs (needs assessment).

Need Assessment is a method to determine the difference between the desired or expected condition (what should be / ought to be) and the existing conditions (what is). Need Assessment methods can be made to measure the level of gaps in students' learning of what was expected and what has been gained. The purposes of this preliminary study are (1) to determine the level of the needs of teachers and the development of the use of assessment to learning that takes place in schools; (2) to design the summative assessment through the PjBL in improving students' higher-order thinking skills.

II. RESEARCH METHODS

A. Participants

In the preliminary study phase, the instrument used is in the form of a needs analysis. The questionnaires were distributed to five teachers of Physics and forty high school students in Bandar Lampung.

B. Procedures

This research applies the research and development method. The research and development method is a research method that produces a particular product and tests the effectiveness of the product [16]. This preliminary study produced summative assessment design through the PjBL in improving higher-order thinking skills of students in high school Physics teaching-learning process. Development procedures performed referring to the study procedures [17] which include the steps of: (1) initial research and data collection; (2) planning; and (3) developing the initial product. The result of this preliminary study is limited to the product of the initial summative assessment design.

C. Assessment Measures

The data collection and needs analysis were carried out by survey method through distribute questionnaires to the teachers and students in five senior high schools in Bandar Lampung. The questionnaires distribution was conducted during the first week in February 2016. The questionnaires were then measured on scale of the answer obtained as the consideration into advanced research. The methods for developing the products were based on literature review and the results of needs analysis.

III. RESULTS AND DISCUSSION

Assessment is an important part in the education. Assessment for learning directly affects student learning by strengthening the relationship between the assessment and instruction. Assessment for learning occurs as a part of the everyday learning processes and uses the information from ongoing assessment to establish learning in the classroom.

Summative assessment is an attempt to summarize learning activities undertaken by a student at a certain time [18]. Summative assessment measures what a student knows at a specific time. It focuses on memorization and recalling skills. Exam grades do not indicate student development. It shows what a student achieved at the end of the course. Multiple choice questions, true-false questions, matching questions, gap-filling (close test) questions, short-answered questions, long answered questions [12]. In the early stages of research, a needs analysis approach is carried out. The results of the analysis of the preliminary study are presented in the following table.

Table 1. Implementation of Assessment by Teacher

No	Analysis Question
1	80% of teachers used assessment to assess students' learning outcomes
2	60% of teachers had not made the assessment in the learning process yet
3	80% of teachers used summative assessment to assess students' learning outcomes
4	80% of teachers agreed that summative assessment through particular learning to assess learning outcomes is developed
5	100% of teachers had difficulty in making the summative assessment to improve students' higher order thinking skills

In Table 1, it can be noted that there are some teachers who has not used summative assessment instrument in assessing the student learning outcomes and had difficulty in making the elements of summative assessment that refers to project-based learning model (PjBL) in improving students' higher-order thinking skills. While the Implementation of assessment in learning according to the students is presented in Table 2.

Table 2. Implementation of Assessment by Students

No	Analysis Question
1	68% Physics teacher made an assessment at the end of each chapter the learning
2	20% of students stated that the assessment made by the Physics teacher only in a written test form
3	88% Physics teachers had the materials selection criteria when assessing learning outcomes
4	80% Physics teachers gave varying matter
5	73% of physics teachers used assessment that cause students to be an active participants and thinking
6	43% of students understood the material they have learned after the assessment carried out
7	78% of students attempted to do physical exercises to develop their own skills
8	65% of students studied if there is a task of the teacher

In Table 2, it is indicated that the students have not really understood the material even though the teachers have been doing variations of materials for the assessment. On the other hand, most students will only repeat their comprehension of learning if there is a task from the teacher. This clearly shows that it takes an appraiser that makes the students able to improve the skills of higher-order thinking owned. Meanwhile, summative assessment needs analysis in measuring the higher-order thinking skills of students by teachers in Table 3.

Table 3. The Needs for Summative Assessment by Teachers

No	Analysis Question
1	80% of teachers would develop a summative assessment to improve students' higher order thinking skills
2	80% of teachers agreed to develop summative assessment to improve students' higher order thinking skills

Based on the analysis in Table 3, it can be concluded that summative assessment is needed to improve students' higher-order thinking skills. While the analysis of summative assessment needed to measure students' higher order thinking skills in students is presented in Table 4.

Table 4. The Needs for Summative Assessment by Students

No	Analysis Question
1	93% of students agreed that teachers of Physics developed a summative assessment (assessment of learning outcomes) through learning which cover the phases of project determination, project design, project implementation, the project evaluation process and results, preparation of reports and presentations, and project completion through monitoring by teachers to improve students' higher-order thinking skills.
2	63% of students agreed that teachers of Physics used summative assessment (assessment of learning outcomes) through learning which cover the phases of project determination, project design, project implementation, the project evaluation process and results, preparation of reports and presentations, and project completion through monitoring by teachers to improve students' higher-order thinking skills.

Based on the analysis in Table 4, it can be concluded that the majority of students agreed in the learning process teachers use summative assessment (assessment of learning outcomes) through learning which cover the phases of project determination, project design, project implementation, the project evaluation process and results, preparation of reports and presentations, and project completion through monitoring by teachers to improve students' higher-order thinking skills.

Based on the analysis of the answers obtained, then the summative assessment is expected to resolve the problems that occurred in increasing higher-order thinking skills of the students. Summative assessment compiled developed through PjBL. Given the summative assessment, it is expected to increase students' ability to think critically by adjusting the needs analysis results obtained through the questionnaire. Several stages done in preparing the design of the summative assessment will be developed through the PjBL in improving students' higher-order thinking skills. The first phase is carried out by preparing the rationale in planning the design of a summative assessment.

This premise is summed up in the form of framework. In the first phase of the scheme, the flow of thinking about the learning process which will be conducted and how the basic concepts assessed by summative assessment is structured. In the concept, it is necessary to distinguish the summative assessment from the formative assessment that has been developed in learning activities at school.

The second stage was reviewing the literature and studies directly to the teacher to determine the extent of the use of assessment in the form of a written test in the study. At this stage, a search through a questionnaire to find out how what students and teachers about the use of summative assessment through the PjBL in learning was conducted. This stage needs support from review of the previous theory that will strengthen the draft of summative assessment design to be conceived and developed.

The third stage is the stage of planning for development. The development which was done was focused on summative assessments through PjBL in improving students' higher-order thinking skills. This summative assessment focused on the project-based learning as the basis. The students were be given a project that ended with the provision of summative assessment in each of the final delivery of learning materials. Thus, the expected higher-order thinking skills that students can increase along with the use of the assessment.

Project-based collaborative inquiry activity provides the greatest support for teachers and students to develop their comprehensive capacity [19]. Project-based learning enable students to learn by doing, engages students' interest and motivates them for learning [20]. Experiential learning and in particular through PjBL, connection with problems of real world is achieved. Students develop apart from cognitive skills, significant abilities that could change our world to a better one, while they enhance their learning outcomes [21].

HOT builds on and extends beyond Bloom's Taxonomy, resulting in discrete dimensions attributed to it: Critical thinking, creative thinking, problem solving, decision making and metacognition, just to name some prominent ones [21]. Based on the research conducted by [14], it is known that there are some aspects that show higher-order thinking skills that a person has namely the ability to think critically, think creatively and solve problems. Critical thinking is an organized process that enables students to evaluate evidence, assumptions, logic, and language that underlies the thinking of others [22]. Besides, critical thinking is enhanced and students can easily conduct their inquiry as well as innovate by exploiting sometimes the advantages of technology [23]. In the project-based learning, by solving different problems it is possible to develop creative ideas while improving highly developed skills [24].

The ability to think creatively includes the ability to create, discover, imagine, predict, design, propose alternatives and produce something. To establish a meaningful creative ideas emerge with something unusual, new, or a solution to a problem. The ability to think creatively can be demonstrated through several indicators, for example, the ability to propose new ideas, ask questions, to make experiment and plan strategies. Critical and creative thinking skill is used in problem solving or trouble shooting. Troubleshooting is transferring the knowledge and skills that exist to answer any unanswered questions or difficult situations. The problem-solving skills is very important because there are always problems in people's lives, including children who are still undergoing formal education in schools.

Learners can encounter problems in learning activities in schools, for example, problems in determining the essay theme, finding a solution to a math problem, or finding material for practicum. The ability to solve problems that a person has can be indicated through several indicators, for example, the ability to identify problems, to have curiosity, to work carefully and to be able to evaluate the decision [14]. The development of students' HOT is complementary with the inculcation of lifelong learning among them. In other words, we need "thinking" students who can incessantly respond to real-world demands [25]. Teachers should play a role to cultivate the use of questions that demand the use of higher order thinking in the classroom while recognizing the difficulties faced by students in solving problems related to HOTS [26].

The higher-order thinking skills critically both the ability of critical thinking, creativity and problem-solving ability of a person cannot be owned directly but rather through practices. The test instrument in form of written tests, besides used to determine the profile of the student's ability, can also be used as a means to train students' abilities to think at a higher level. The test items to use can contain questions that test students in problem solving, critical thinking and creative thinking. In order to answer the questions, the students need higher-order reasoning which is high logical way of thinking. Thinking high logically is needed by students in the learning process in the classroom, especially in answering questions because the students need to use the knowledge, comprehension and skills they have and relate them in new situations.

In preparing the summative assessment instruments, the indicators of three aspects of higher-order thinking skills, the ability of critical thinking, creative thinking, and problem solving are developed [27]. The aspects of critical thinking skills consist of 6 indicators which are the ability to ask questions, to revise the wrong concept, to plan strategies, to evaluate the decision, to criticize a statement, and to evaluate the decision.

The aspects of creative thinking ability is composed of 12 indicators that are students ability to formulate equations, make connections between concepts, propose new ideas, prepare relationship of the concepts in schematic form, illustrate ideas, be dare to make experiments,

organize concepts, produce something new, design experiments, modify concepts with new things, combine the concept of a coherent and to change the equation. The test items to test creative thinking abilities of students to solve problems are in the form of images and present problems that can bring the creativity of students.

Aspects of problem-solving ability is composed of 11 indicators that cover students ability to identify problems, state a causal connection, apply the concept in accordance with the problem, have curiosity, create charts or images to solve a problem, explain some possibilities as a solution, be open-minded, make decisions, work carefully, speculate and reflect on the effectiveness of the problem solving process.

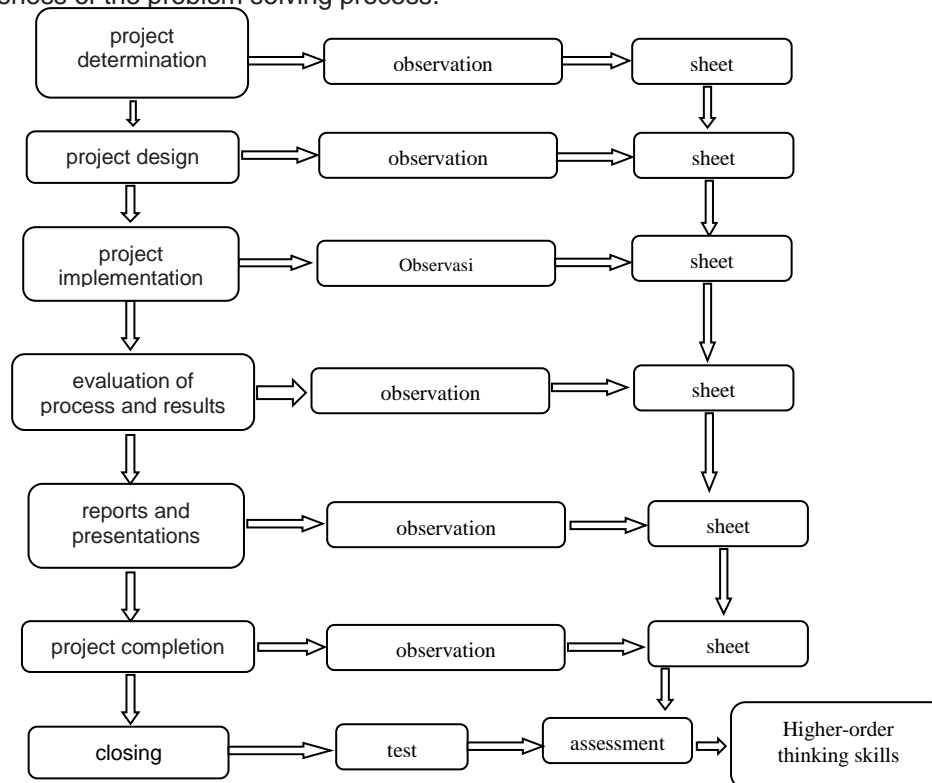


Figure 1. Summative Assessment Design through the PjBL to improve Student's Higher-Order Thinking Skills

The summative assessment to improve higher-order thinking skills is conducted at each phase of PjBL. The phases to perform are the project determination, project design, project implementation, the project evaluation of process and results, preparation of reports and presentations, and project completion through monitoring by teachers.

1. Project determination

In this step, the students determine the theme or topic of the project based on project assignments given by the teacher. Students can develop problems as the determination of the initial framework of the project. Students are given the opportunity to select or specify the projects to be doing well in groups or independently with the record does not deviate from the assignment of teachers.

2. Project design

Students devise steps to completion of the project from start to finish and their management. This project designing activity contains rules in the implementation of the project tasks, the selection of activities to support the work of the project, the integration of a variety of possible completion of project tasks, resources planning or materials or tools that can support the completion of project tasks, and cooperation among members of the group.

3. Project implementation

Students under the mentoring of teachers to schedule all the activities that have been designed. How long the project should be solved step by step.

4. The project evaluation of process and results

This step is creation of project design implementation. Activities that can be done in the project activities of which is to a) read, b) researching, c) observation, d) interviews, e) record, f) works of art, g) visited the project object, or h) internet access. Teachers responsible for monitoring students' activity in the conduct of the project start process until the completion of the project. In the monitoring activity, the teacher made rubric that will be able to record the activity of students in completing project tasks.

5. Preparation of reports and presentations

The results of the project in the form of a product, whether it be paper products, artwork or works of technology or craft presented and or published to other students and teachers or the community in the form of exhibitions learning products.

6. Project completion through monitoring by teachers

Teachers and students at the end of the learning process of reflection on the activities and results of project tasks. The process of reflection on project tasks can be done individually or in groups. In the evaluation phase, students are given the opportunity to complete the task bring his experience during the project that developed the discussion to improve performance for completing project tasks. At this stage also carried feedback on processes and products that have been produced.

7. Closing

It is the final step in learning. Students are given about the form of summative assessment to gauge the understanding gained during learning. Problem is also high-order thinking skills possessed by students. The teacher observe the students' activity as long as project based learning process, from the first step to the sixth step. The observation results of project-based learning implementation is showed by student projects activity sheet.

IV. CONCLUSION

Based on the results and discussion, it can be concluded that (1) the teacher is relatively rare in carrying out the teaching through project-based learning; (2) it is necessary to have an assessment that can enhance students' higher-order thinking skills through the PjBL.

The summative assessment to improve higher-order thinking skills is conducted at each phase of PjBL. The phases to perform are the project determination, project design, project implementation, the project evaluation of process and results, preparation of reports and presentations, and project completion through monitoring by teachers.

ACKNOWLEDGMENT

This project would not be possible without the help from member of our team who are not authors on this paper Tuti Widyawati, Astri Mela Agustin, Muhammad Iwan, Novinta Nurulsari, Nurul Ulil Amri, Nur Hasanah, and Refdi Akmal.

REFERENCES

- [1] G.S. Hanna and P.A. Dettmer, *Assessment for effective teaching: Using context-adaptive planning*, Boston, MA: Pearson A&B, 2004.
- [2] K. Earl and D. Giles, "An-Other Look at Assessment: Assessment in Learning, New Zealand: New Zealand Journal of Teacher's Work", 2011, pp. 11-20.
- [3] E.R Lai, *Performance-based Assessment: Some New Thought on Old Idea*, Bulletin Issue 20, 2011, pp. 1-4. Retrieved from Pearson website: <http://researchnetwork.pearson.com/wp-content/uploads/Performance-based-assessment.pdf>
- [4] NSW, *Assessment For As and Of Learning*, BOSTES, 2016. Retrieved April 20, 2016, from New NSW Syllabuses: <https://syllabus.bostes.nsw.edu.au/support-materials/assessment-for-as-and-of-learning/>
- [5] L. Nevin, *Technology and Differentiated Instruction*, 2016. Retrieved from <http://www.tvdsb.ca/webpages/takahashid/techdia.cfm?subpage=128207>

-
- [6] Kerry J. Kennedy, J.K.S. Chan, P.K. Fok, and .M. Yu, "Form of assessment and their potential for enhancing learning: conceptual and cultural issues, Hong Kong: Educational Research For Policy and Practice vol. 7 number 3, 2008, pp. 197-207.
 - [7] Maddalena Taras, "Assessment for learning: assessing the theory and evidence", UK: Procedia Social Behavioral Sciences vol. 2, 2010, pp. 3015-3022.
 - [8] D. Wiliam, What is assessment for learning?, UK: Studies in Educational Evaluation vol. 7(1), 2011, pp. 3-14.
 - [9] D.R. Sadler, "Formative assessment and the design of instructional strategies", Instructional Science vol. 18, 1989, pp. 119-144.
 - [10] H. Karal and Z. Celik, Designing components of a sample system considering effect of web-based technologies on meaningful measurement and assessment practices, Turkey: Procedia Social Behavioral and Sciences vol 9., 2010, pp. 908-912.
 - [11] J.David, What Research Says About/Project-Based Learning, California: Educational Leadership Teaching Students to Think vol. 65 number 5, 2008, pp. 80-82.
 - [12] Gamze Sart, (2014). "The effects of the development of metacognition on project-based learning". Turkey: Procedia - Social and Behavioral Sciences vol. 152, 2014, pp. 131-136.
 - [13] Abdul Malik, Chandra Ertikanto, and Agus Suyatna, "Description the needs of HOTS assessment on Physics learning with guided inquiry method", Transl., Indonesia: Prosiding Seminar Nasional Fisika, 2015, pp. 1-4.
 - [14] Emi Rofiah, Nonoh Siti Aminah, Elvi Yusliana Ekawati, "The preparation of Physics higher-order thinking skills test instrument on junior high school", Indonesia: Jurnal Pendidikan Fisika vol.1, 2013, pp. 17-22.
 - [15] Yee Mei Heong, Jailani Md Yunos, Widad Othman, Razali Hassan, Tee Tze Kiong, Mimi Mohaffyza Mohammad, "The needs analysis of learning higher order thinking skills for generating ideas", Malaysia: Procedia - Social and Behavioral Sciences vol. 59, pp. 197-203.
 - [16] Sugiyono, Method of Quantitative, Qualitative, Research and Development", Bandung: Alfabeta, 2011.
 - [17] W.R. Borg, J.P. Gall, and M.D. Gall, Educational Research: An Introduction, Boston, MA: Longman, 2003
 - [18] D. Kennedy, "Writing and Using Learning Outcomes Writing and Using Learning Outcomes – A Practical Guide Practical Guide, Quality Promotion Unit. Irlandia: University College Cork, 2007.
 - [19] Shaoqing Guo and Yanjun Yang, "Project-based learning: an effective approach to link teacher professional development and students learning", Northwest: Journal of Educational Technology Development and Exchange vol. 5 number 2, pp. 41-56.
 - [20] Utku Kose, "A web based system for project-based learning activities in 'web design and programming' course", Turkey: Procedia Social Behavioral Sciences vol. 2, 2010, pp. 1174-1184.
 - [21] Tan Shin Yen and Siti Hajar Halili, "Effective teaching of higher-order thinking (HOT) in education", The Online Journal of Distance Education and e-Learning vol. 3 Issue 2, 2015, pp. 41-47.
 - [22] E. Johnson, Contextual Teaching & Learning, Bandung: MLC, 2007.
 - [23] Larmer, J. ,Mergendoller, J. R. (2010).Seven Essentials for Project-Based Learning. Educational Leadership, 68, 1, 34-37.
 - [24] Douladeli Efstratia, "Experiential education through project based learning", Greece: Procedia – Social Behavioral Sciences vol. 152, 2014, pp. 1256-1260.
 - [25] Phawani Vijayaratnam, "Developing higher order thinking skills and team commitment via group problem solving: a bridge to the real world", Malaysia: Procedia – Social Behavioral Sciences vol. 66, 2012, pp. 53-63.
 - [26] Abdul Halim Abdullah, Nur Liyana Zainal Abidin, and Martina Ali, "Analysis of students' errors in solving higher order thinking skills (HOTS) problems for the topic of fraction", Canadian: Canadian Center of Science and Education vol. 11, 2015, pp. 133-142.
 - [27] N. Mourtos, N. Okamoto, and J. Rhee, "Defining, teaching, and assessing problem solving skills", India: 7th UICEE Annual Conference on Engineering Education, 2004, pp. 9-13.

Assessment Model Multiple Intelligences Learning Approach in Primary School Mathematics Subjects

⁽¹⁾Helmiyah Suryani, ⁽²⁾Badrun Kartowagiran

⁽¹⁾Dinas Pendidikan Kota Samarinda, Jalan Untung Surapati No 1, Telpon 0541

⁽²⁾Universitas negeri Yogyakarta, Jalan Karang Malang 55281 Yogyakarta

Email: emysuryani29@yahoo.com

badrunkw@yahoo.com/ kartowagiran@uny.ac.id

Abstract-This research is intended to develop learning assessment with multiple intelligence approach. This model is developed to be used by teachers that have been using multiple intelligence in their teaching (kbn). The method in this research is using research and development approach by Borg and Gall. The development stages into three: pre development stage, development stage, and method application. Predevelopment was an introductory study to Mutiara Ilmu Elementary School at Pasuruan, and Yami Elementary school East Java, literature review. The development is made of the making of multiple intelligence indicator, construct validity by five experts, the making of curriculum indicator on second year of elementary school students, the making of test items, instrument seminar, and fgd by seven experts. Model application is made of the trial, field test, final revision, and dissemination. The object of this study is 221 students of second grader at Muhammadiyah 1 Elementary school Samarinda. Instrument is 40 items in summative model multiple choices [form](#). The trial was conducted twice, which was trial and field test. Multiple intelligence's construct validation was done by five experts. The Aiken analysis showed that two indicators were not usable. Thus only 34 indicators were valid ranging from 0.70 to 1.00. The earlier model that was made of 64 items validated by seven [experts](#). The Aiken analysis showed from those 64 items, 49 are valid ranging from 0.74 to 0.9. The trial took 40 items. The trial results and field test were analysed with Quest. Research Results: 1. The instrument that has been developed is 37 items of multiple choice. 2. Quest analysis has resulted 37 fit items within vertical lines or having infit MNSQ between 0.81 until 1.30. The instrument has infit MNSQ average of 1.00 with standard deviation of 0.10. The reliability of the instrument is 0.78. This research has succeeded to find the assessment model with multiple intelligence approach that is made of procedures of use, the assessment instrument of learning with multiple intelligence approach, and scoring guide.

Keywords: *Multiple Intelligences, Learning Assessment*

I. INTRODUCTION

In the National Standard for Education (Permendiknas No. 66, year 2003) stated that from all eight national standard for education, one of them is assessment standard is intended to guarantee:

- a. The planning of assessment of learning participant in accordance with competencies that are aimed and based on assessment principles:
- b. professional, open, educative, effective, efficient, and social cultural context relevant assessment execution;
- c. objective, accountable, and informative result report. This Standard Educational Assessment is arranged as an assessment guide for educator, education force, and government in educational force for elementary and middle high school.

This Multiple Intelligence Assessment or Assessment with Multiple Intelligence approach is an assessment that follows the learning process that uses multiple intelligence approach. This assessment is only applied when the learning is done using multiple intelligence approach. The theory that is used to develop this assessment is Multiple Intelligence approach founded by Howard Gardner. According to Howard Gardner human's intelligence is grouped to: linguistic

intelligence, logical mathematic, spatial intelligence, musical intelligence, kinesthetic intelligence, interpersonal intelligence, intrapersonal intelligence, naturalist, and existentialist.

The background of this research development: preliminary study report on Mutiara Ilmu Pasuruan Elementary School, YAMI Gresik Elementary School, and Muhammadiyah 1 Samarinda Elementary School. On those three schools, multiple intelligence approach in learning is applied, while the assessment on multiple intelligence approach is yet to be used.

A. The problem found is there is no match between the learning process approach and the assessment process approach.

The research about human potential through "Harvard Project Zero" that has found eight human intelligences that later will be called as multiple intelligences. In a few journals in America, it is known that the scheme implemented to the theory in multiple intelligence can improve teacher's ability in developing creativity in new and interesting teaching process. Multiple Intelligence's learning process is also proved to be able to increase student's achievement, this is proved by several research. Gardner (1983:16) said that multiple intelligence is influenced by the culture we are born into so it can no longer be translated as a singular word in cognitive field.

The purpose of this research is to find learning assessment model with multiple intelligence approach. This model can be used by teachers that have used multiple intelligence's model as their learning delivery approach. In the development process, researcher has limited to only three multiple intelligence facets, those are logical mathematical intelligence, spatial intelligence, and linguistic intelligence. The assessment model is developed based on Multiple Intelligence's theory by Howard Gardner (Gardner, 2003) that grouped intelligences into eight clusters:

1. Linguistic Intelligence
2. Logical Mathematical Intelligence
3. Kinesthetic Intelligence
4. Musical Intelligence
5. Spatial Intelligence
6. Interpersonal Intelligence
7. Intrapersonal Intelligence
8. Naturalist Intelligence
9. Linguistic Intelligence: sensitivity to sound, structured, interpretation, word's function in language. Related to writing skill, reading, making discussion, arguing.
10. Logical Mathematical Intelligence: logical pattern and numerical sensitivity, able to process long flow of thought. Related to counting skill, reasoning, and logical thinking in problem solving.
11. Visual-Spatial Intelligence: sensitivity in feeling and imagining visual and spatial world accurately. Related to drawing skill, taking picture, sculpting, and designing.
12. Musical Intelligence: sensitivity and ability to create and appreciate rhythm, notes pattern, and forms of emotional expression in music. Relate to the ability to create song, listening notes/sound from instruments.
13. Kinesthetic Intelligence: The ability to control body movement and expertise in tinkering with objects, response, and reflects. Related to motoric movement and body balance.
14. Interpersonal Intelligence: Sensitivity to digest and respond accurately to mood, temperament, motivation, and other people's will. Related to the ability to mingle, to lead, social sensitivity, negotiation, cooperation, and highly empathetic.
15. Intrapersonal skill: Sensitivity to one's own feelings and able to differentiate emotion, knowledge about one's own strengths and weaknesses. Related to abilities to control oneself, intuitive ability, and self motivation.
16. Naturalist Intelligence: An ability to know and differentiate species' members, to acknowledge the existence of other species, and to map the relationship between several species. Related to the ability to research natural symptoms, to classify, and to identify.

B. The characteristics of those eight multiple intelligences are:

In developing this assessment model, researcher limited the scope of this research to only three intelligences; logical mathematical intelligence, visual-spatial intelligence, and linguistic intelligence. This is decided to limit the width and depth of this research material developed (Campbell, et al:2002: 10).

Munif chatib (2009: 102) said that "one's intelligence is developing, not static. One's intelligence is highly related with habit, repeated behavior". Conducting learning process with multiple intelligence approach and followed by the same type of assessment model is a pattern to form repeated behavior in student's intelligence development.

Multiple Intelligence Assessment Model is developed on math because school's mathematic is learning a basic material and is a pre requisite to move to a higher concept. The concept learned is easily applicable to daily life, and can be approached by student's experience with real stuff around them.

This research is important to get (1) multiple intelligence assessment model that later will be used against learning process with multiple intelligence approach, (2) to motivate teachers to develop multiple intelligence assessment model, (3) to help teachers to design learning and assessment.

II. METHODS

This research is using Borg & Call research and development approach, which modified into three stages: pre development stage, developmental stage, and model application. The pre development stage consists of; preliminary study in Mutiara Ilmu Pasuruan Elementary School in East Java, YAMI Elementary School in Gresik, Muhammadiyah 1 Elementary School in Samarinda, East Kalimantan, and literature review. The developmental stage consists of: the making of multiple intelligence's indicator, construct validity by five experts, the making of mathematic curriculum's indicator for second grade of elementary school students, the making of items, instrument seminar, and Focused Group Discussion by seven experts. The Model Application stage consists of trial, field test, final revision, and dissemination. The population of this research is second grader of elementary school. The sample came from 112 students of second grader in elementary school. The instrument is in forty items of multiple choices form.

III. RESULTS AND DISCUSSION

A. Construct Validity

The indicators of multiple intelligence from mathematical-logical intelligence, visual-spatial, and linguistic are arranged based on its theory. Arranged in 37 indicators, after construct validity is conducted by five experts, and analysed with Aiken analysis, they have resulted in range from 0.60 until 1.00. Three indicators failed due to being in the validity of 0.60 which are indicator no. 15, 29, and 33. Other indicators are ranging from 0.73 -1.00 as as many as 34 indicators are made through construct validity.

B. Content Validity

Early model of draft are made by valid indicators. Every indicator is reflected by two items. 64 items in early model of draft which next to be content validated by three ways: (1) instrument seminar in which attended by doctoral students, (2) validated by seven experts, 1 measurement expert, 1 educational evaluation expert, 2 mathematical expert, 1 psychometric expert, and 2 psychological expert. Experts' content validity later analysed by Aiken analysis. The result of this validity is the fall of 8 items that are lower than 0.7, which are no. 3, 12, 13, 21,22,37,43, and 44, all these items are fallen. The other 56 items is ranging from 0.71 – 0.92. From all 56 items that are valid, the package of for trial which consists of top 40 items is arranged.

Trial the early model which consists of 40 items are next to be tested against 69 second grader of Muhammadiyah 1 Elementary School Samarinda. The result of this test is analysed with IRT using QUEST and ITEMAN for classical theory test.

Test result by IRT (QUEST)

Items' characteristic: Level of difficulty between -1.65 – 1.44

Items fell: 19 since > 2 (2.48)

Distinguishing power: There are 38 items that are ranging from 0.00 – 0.58

Two items fell due to its negative result, which are no. 20 (-0.02) and no. 31 (-0.04)

Reliability: 0.78

Table 1. Analysisi by IRT (Quest)

MULTIPLE INTELLIGENCE DIMENSION	Indicator	Item	Thresholds	Pt Biserial	Reliability
1.Logical Mathematical	1.1	8 14	0.88 -1.65	0.38 0.06	0.78
	1.2	12	-0.13	0.54	
	1.3	1 31	-0.92 -0.67	0.35 -0.04	
	1.4	5	-0.28	0.00	
	1.5	21 27	-0.067 1.44	0.37 0.23	
	1.6	6	-0.67	0.03	
	1.7	26	-0.46	0.14	
	1.8	28	-0.13	0.21	
	1.9	2 3	-0.46 -1.23	0.30. 21	
	1.10	24 39	1..21 -0.67	0.19 0.33	
	1.11	34 37	-0.67 0.50	0.11 0.34	
	1.12	15	-0.46	0.44	
	1.13	11 20	-0.46 -1.65	0.37 -0.02	
	1.15	9 10	0.50 1.13	0.42 0.30	
2.Visual-Spatial	2.1	7	-1.23	0.04	
	2.2				
	2.3				
	2.4				
	2.5				
	2.6	29	-0.46	0.50	
	2.7				
	2.8	32	0.97	0.39	
	2.9	4, 29	-0.46 -0.46	0.27 0.50	
	2.10	16, 17	-0.46 1.13	0.44 0.29	
	2.11	13, 35	-0.28	0.49 0.00	
3. Linguistic	3.1	18, 38	1.37 -0.28	0.33 0.35	
	3.2	19	2.48	0.27	
	3.3				
	3.4				
	3.5	33, 36	-0.13 -0.13	0.32 0.11	
	3.6	22, 30	0.02 0.39	0.57 0.33	
	3.7	23, 40	0.60	0.58	

Analysied by ITEMAN

Probability ranging between 0,435 – 1

Biserial Point between 0.020 – 0,580

One item fell due to its negative result which is item no. 35 (-9.00)

Reliability at : 0,742

Table 2. Trial Analysis' Test Results Clasical Theory (ITEMAN)

DIMENSI KECERDASAN MAJEMUK	INDIKATOR	BUTIR	Probabilitas Correct	Pt Biserial (Daya Beda)	Reliabilitas
1. Logis Matematis	1.1	8 14	0.75 0.971	0.38 0.07	0.74
	1.2	12	0.88	0.55	
	1.3	1 31	0.94 0.93	0.35 -0.04	
	1.4	5	0.99	0.22	
	1.5	21 27	0.93 0.66	0.38 0.23	
	1.6	6	0.93	0.03	
	1.7	26	0.91	0.14	
	1.8	28	0.89	0.46	
	1.9	2 3	0.91 0.96	0.33 0.21	
	1.10	24 39	0.70 0.93	0.19 0.33	
	1.11	34 37	0.93 0.81	0.11 0.34	
	1.12	15	0.91	0.45	
	1.13	11 20	0.91 0.98	0.37 -0.02	
	1.15	9 10	0.81 0.71	0.42 0.30	
2. spasial	2.1	7	0.96	0.04	
	2.2				
	2.3				
	2.4				
	2.5				
	2.6	29	0.91	0.51	
	2.7				
	2.8	32	0.74	0.40	
	2.9	4, 29	0.91 0.91	0.27 0.51	
	2.10	16, 17	0.71 0.71	0.36 0.30	
	2.11	13, 35	0.90 1.00	0.50 -9.00	
3. Linguistik	3.1	18, 38	0.67 0.90	0.33 0.56	
	3.2	19	0.435	0.27	
	3.3				
	3.4				
	3.5	33, 36	0.88 0.88	0.33 0.11	
	3.6	22, 30	0.87 0.83	0.57 0.33	
	3.7	23, 40	0.80 0.84	0.58 0.35	

Findings in this research which have resulted from its trial has made 37 items consists of 20 items of logical mathematical intelligence dimension, 8 items of visual-spatial intelligence dimension and 9 items of linguistic intelligence dimension that have fulfilled the requisite of validity and reliability.

Table 3. Items That are Valid & Reliable

NO	Dimension	No. of Items	Difficulty Level	Distinguishing Power
1	Logical Mathematical	20	-1,65 - 1,44	0,00 – 0,58
2	Visual-Spatial	8		
3	Linguistic	9		
Total		37		

IV. CONCLUSION AND SUGGESTIAN

This Multiple Intelligence Assesment Model has fulfilled the requisite of validity and reliability to be used in the assessment of learning with multiple intelligence approach.

Suggestion 1: This multiple intelligence model need to be developed to other intelligence than these three that have been developed in tis research. 2. The Assessment Model of Multiple Intelligence still needs to be developed for elementary school other than second grader, to middle high school, or high school.

REFERENCE

- [1] Anderson, L. W & Krathwoh , D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Blomm's taxonomy of Educational objective*. New York:Longman.
- [2] Armstrong, T, (2003). *Sekolah Para Juara*. Bandung: Kaifa.
- [3] Allen,M.J &Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole).
- [4] Badan Standarisasi nasional Pendidikan (2006).
- [5] Borg, W.R., & Gall, M. D. (1983). *Educational research: An introduction*. London: Longman, Inc.
- [6] Munif Chatib. (2009). *Sekolahnya manusia*. Bandung: Kaifa.
- [7] _____ (2012). *Sekolah anak-anak juara*. Bandung: Kaifa.
- [8] Campbell, L & Campbell, B & Dickinson, D. (2002). *Teaching & learning through multiple Intelligences (Metode Terbaru melesatkan kecerdasan)*. Inisiasi Press.
- [9] Chan, D.W.(2004). *Multiple Intelligences of Chines gifted students in hongkong: Perspective from Students, Parent, Teacher, and Peers*. 18/Roeper Review, vol. 27 N0.1. 18-24
- [10] Djemari Mardapi. (2004). *Penyusunan tes hasil belajar*. Yogyakarta . Program Pascasarjana Universitas Negeri Yogyakarta.
- [11] _____ (2007). *Teknik Penyusunan instrumen tes dan non tes*. Yogyakarta: Mitra Cendikia Press .
- [12] Dryden, G. & Jeannette, V. (2003). *The Learning revolution (Revolusi cara belajar)*. Bandung.:Kaifa.
- [13] DePorter, B. & Reardon, M. & Singer-N., (2014). *Quantum teaching*. Bandung: Kaifa.
- [14] Gardner, H. (1983). *Frame of Mind*. BasicBooks.
- [15] Gardner, H. (2003). *Multiple inteelligences*.Interaksara.
- [16] Gardner, H. (1999). *Intelligence Reframe. Multiple Intelligence for The 21 Century*. Basic Books.Stiggins, R.J & Chappuis, J.(2011). *An Introduction to student –involved assessment for learning*, US: Pearson.
- [17] Retnawati, H. (2014). *Teori Respon Butir dan penerapannya*. Nuha Medika.
- [18] Retnawati, H. (2016). *Validitas Reliabilitas & Karakteristik Butir*. Nuha medika.

INDICATOR DEVELOPMENT OF LEARNING MODEL EVALUATION INSTRUMENT

Herpratiwi¹, Tien Yulianti², Adil Fadlilah H³, Bajawati⁴

¹Lecturer of Lampung University, ²Lecturer of STBA Teknokrat Lampung,

³Supervisor of Madrasah M.Ts/ MA Bandar Lampung,

⁴Teacher of SMA 13 Bandar Lampung

E-mail address: herpratiwi64@yahoo.com

Abstract—Professional educators are required to develop a learning model that was based on a constructivist understanding. This study aimed to obtain the indicators needed to measure a learning model developed by educators in all types and levels of education. Indicator is intended to evaluate the components of a learning model.

The method used in this research is quantitative descriptive. Data was collected by the Focus Group Discussion (FGD) with three steps; a preliminary investigation, design, and realization / construction. It was involving six experts and four practitioners. Validation of the construct is in the form of expert judgment. The opinions of experts and practitioners were analyzed descriptively, while the reliability of the instrument was analyzed by Cronbach's Alpha.

The results showed that there were 7 indicators of learning model developed from five aspects; rational theoretical (2 indicators and 11-item criteria), syntax (1 indicator and three item criteria), the principle of interaction (1 indicator and 7 item criteria), social systems (1 indicator and three item criteria) and the impact of learning (2 indicators and criteria item 7). The measurement results were using 3-scale models of learning; 1 (= if the indicator is less good / less logical), 2 (= if the indicator is good enough / illogical) and 3 (= if the indicator is good / logical). Assessment categories were grouped into three; No Good (TB), Good (B) and Very Good (SB). Reliability value questionnaires are rational theoretical aspects of 0.89, the syntax of 0.81, 0.93 interaction principle, the social system and the impact of learning 0.86 0.84. Based on reliability testing all aspects of the value of Cronbach's Alpha of 0.94, thus all items contained in the questionnaire are reliable and all indicators internally consistent because they have a strong reliability.

Key words: five aspects, learning model, seven indicators

I. INTRODUCTION

Professional teachers in all types and levels of education are required to have professional competence, pedagogical, social and personality. One of the accomplishment in the learning process, teachers are required to provide education in an interactive, inspiring, fun, challenging, motivating the students to actively participate and provide enough space for innovation, creativity, and independence in accordance with their talents, interests, and physical development as well as psychological learners

Moreover, in the process of learning, educators should prioritize the planting of the character values in addition to the transfer of knowledge and skills. Therefore, educators must design the advanced needs of the students, as the outlined in lesson planning. Instructional design promotes character education and encourages students to think high to achieve learning goals effectively and efficiently, if the design is embodied in the learning model. Educators can design their own model of learning and use learning model that has been designed by experts or other practitioners.

According to Arends (2008) learning model has four characteristics; (1) rational theoretical, has a foundation to think how the nature of learners can study well, (2) learning objectives,

namely what is the purpose of students learning, (3) the syntax, that is how sequence patterns of behavior of learners with educators, and (4) how to support the learning environment. Meanwhile, according to Bruce Joyce and Weil (2009), learning model is as a conceptual framework that describes a systematic procedure for organizing the learning experience of students.

In detail, Bruce Joyce and Weil (2009), explains the components of learning models, namely: (1) rational theoretical, as a basis to think how the nature of learners can study well, (2) syntax, how the sequence pattern of behavior of learners with educators, (3) the principle of interaction, how educators positioned themselves against educators and learning resources, (4) social system, how do you view among components within a social community, (5) support system, how a supportive learning environment, and (6) the impact learning, namely how the results and impact of the expected good learning instructional impact (instructional effect) and the impact of Bridesmaids (nurturant effect).

Since it is lack of an evaluation tool that can be used by teachers to measure the design of learning model by themselves or other experts, it needs to develop the indicators of evaluation instrument. So the evaluation instrument to measure a learning model is available.

The problem of this study is how the development of the indicator of learning model can provide information about a model of learning. The objective of the study is to get the instruments indicators on learning model. The advantage of the study is teachers can use a set of indicators to evaluate a learning model which is designed by themselves or other experts.

II. THE METHODS OF STUDY

This type of study is development research. The steps are preliminary investigation, design, realization/construction, test, evaluation and revision, and implementation (Plomp, 1977). The steps are adjusted with the development of a learning model evaluating indicators, it is only up to the stage three.

The first stage is introduction; preliminary study activity and collection of information are taken due to the learning model that is often used by teachers. It is intended to seek information about the weaknesses and strengths of the learning model. It is also to examine theories of learning models, as well as the assessment of the relevancies of previous studies. Second step is the design phase; the development of components and aspects of learning and the planning model evaluation base on the grid instruments of data collection. The third step is realization/construction phase; the construction is validated by expert in order to know whether the indicators of learning model evaluation instruments appropriate to the theory/construction.

This activity is done by 6 experts in Focus Group Discussion (FGD) to analyze if the indicator was suitable to the construction. In this step, Review of practitioners is performed by 4 senior teachers. After FGD and review were done, the indicators were revised to repair and to take other inputs Cronbach's Alpha.

III. THE RESULT AND THE DISCUSSION OF STUDY

Based on the result of FGD of experts and practitioners, it is released the indicators of learning model evaluation instruments to be used to measure a learning model. The result of the study is available on Table 1.

Table 1. The Result of Indicator Development of Learning Model Instrument.

No.	Aspect	Indicator	Criteria
A.	Rational Theoretical	There are theories relate to the character of subjects	The dimension clarity of cognitive process to be achieved
			The dimension clarity of knowledge to be achieved
			The clarity of competency level to be achieved
			The clarity of affective domain level to be achieved
			The clarity of affective domain type to be achieved
			The clarity of psychomotor domain level

No.	Aspect	Indicator	Criteria
		There are theories relate to the character of students	to be achieved
			The clarity of learners' early competency
			The clarity of learners' learning motivation
			The clarity of learners' interest
			The clarity of learners' socioeconomic background
B.	Syntax	The sequence of learning steps that must be done by teachers when going to use certain learning model	The clarity of learners' multiple intelligence
			The clarity of the sequence of learning steps that must be done by teachers
			The clarity of the approaches, strategies, methods, techniques and tactics that available in a model
C.	Principle of Interaction	The patterns of interaction both learners and teachers in the learning and assessment	The clarity of both hierarchical sequence of steps or procedures
			There is educational interaction between learners and educators
			Educational interaction between learners and learners
			Variation of interaction between learners and educators
			Variation of interaction between learners and learners
			The interaction between students and other learning resources, which are designed and utilized
			The frequency or more or less the turn of the action between teachers and learners
D.	Social System	The design of collaboration between learners and teachers in learning	The frequency or more or less the turn of the action between learners and learners
			Learners' participation in individual learning
			Learners' participation in group learning
E.	Instructional Effect	The effect of direct learning (Instructional Effect)	Teachers' participation in learning
			The increasing of learning achievement of cognitive process dimension
		The effect of indirect learning/supporting (Nurturing Effect)	The increasing of learning achievement of psychomotor dimension
			The increasing of soft skill of value affective dimension
			The increasing of soft skill of motivation affective dimension
			The increasing of soft skill of value affective manner
			The increasing of soft skill of value affective emotion
			The increasing of soft skill of value affective interest

Based on the analysis of experts and practitioners, to evaluate a learning model there are five aspects and 7 measurement indicators, and developed criteria as a measure of each indicator. Table 2 shows the number of indicators and criteria from each aspects.

Table 2. The Number of Indicators and Criteria of Learning Model

No.	Aspects	Amount of Indicators	Total of Criterion
1.	Rational Theoretical	2	11
2.	Syntax	1	3
3.	The Principle of Interaction	1	7
4.	Social System	1	3
5.	Instructional Effect	2	7
Total		7	31

The result of learning model measurement uses 3 scales; 1 (=if the indicator is not good/illogical), 2 (=if the indicator is enough/illogical), and 3 (=if the indicator is good/logic). Assessment categories are grouped into three, they are No Good (TB), Good (B) and Very Good (SB). The way of calculating the category follows the opinion of Arikunto (2009), as showed on Table 3.

Table 3. Assessment Categories

No.	Reference	Category
1.	$\geq (\bar{x} + 1. \text{SD})$	Very Good (SB)
2.	$(\bar{x} - 1. \text{SD}) \leq \sum < (\bar{x} + 1. \text{SD})$	Good (B)
3.	$< (\bar{x} - 1. \text{SD})$	Not Good (TB)

Table 4 shows that the value of reliability questionnaires rational theoretical aspects of 0.89, the syntax of 0.81, 0.93 interaction principle, the social system and the impact of learning 0.86 0.84. Based on reliability testing all aspects of the value of Cronbach's Alpha of 0.94, with demikism all items contained in the questionnaire is reliable and all indicators internally consistent because it has strong reliability, (Maier, U., Wolf, N., & Randler, C. , 2016; Bonett, DG, & Wright, TA, 2015; Sebastian Rainsch, 2004).

Table 4. The reliability Indicator Score of Learning model

No.	Aspek	Nilai Cronbach's Alpha
1.	Rational Theoretical	0,89
2.	Syntax	0,81
3.	The Principle of Interaction	0,93
4.	Social System	0,86
5.	Instructional Effect	0,84
6.	All aspects	0,94

IV. CONCLUSION

The conclusion of this study is the indicator development of learning model instruments through expert judgment, FGD, and practitioners' review can be used to measure a learning model. Since indicator is based on the clear construction, so it is possible to expose a learning model.

REFERENCES

- [1] Arens, Richard. (1998). *Learning to Teach (International Edition)*. Singapore : Mc. Graw Hill.
- [2] Arikunto, Suharsimi. (2009). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- [3] Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, 36(1), 3-15.
- [4] Bruce Joice & Weil Marsha. (2009). *Models of Teaching*. USA: Perason Education.
- [5] Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, 95, 85-98.
- [6] Plomp. T. (1997). *Development Research in/on Educational Development*. Netherlands: University of Twente.
- [7] Sebastian Rainsch. (2004). *Dynamic Strategic Analysis: Demistyfying Simple Succes Strategies*. Wiesbaden: Deutscher Universitasts-verlag.

PERFORMANCE ASSESSMENT IN MODEL OF LEARNING SUPERFLEX®

Huriah Rachmah

Magister Program of Social Studies STKIP Pasundan

huriahrachmah@gmail.com

Abstract-- Superflex® learning model to teach students who have difficulties in socializing with his friends at the school in solving problems. Learning aims to help students use strategies to organize itself, become a social thinker and have social skills. Superflex® curriculum is developed based on the theory of cognitive behavioral approaches. Cognitive behavioral theory aims to increase one's awareness about the impact of the behavior of students in others and ourselves. Assessment is carried out in this model is the performance appraisal. Assessment of performance as an authentic assessment will observe the behavior of the students and the performance review process on the situation. This assessment can measure the ability of social thinking and social skills of students so the success of this model can be measured.

Keyword: Performance Assesment, Superflex®, Model of Learning

I. INTRODUCTION

Model is a comprehensive planning approach to help students learn all kinds of knowledge, attitudes and specific skills. The learning model has a theoretical base that includes learning steps necessary to reach the expected educational outcomes (Joyce, Weil, & Calhoun, 2009). According Astaty (2010) in Rachmah (2013) learning model is an activity that illustrates the process from start to finish and was presented with the approaches, methods, and specific learning techniques.

Superflex® learning model was developed based on the theory of cognitive behavioral approaches to explore how social thinking. According to Attwood (2006) in Rachmah (2015) Cognitive behavioral goal to increase one's awareness of the impact of his behavior towards others and oneself. Model Superflex® learning can help students to increase self-awareness of behavior and learn how to modify negative behavior by using super flexible strategy.

Based on the results of research conducted by Winner (2014) inability problem learners for social thought can be trained with Superflex® learning model. Learners are invited to become a social detective who can perform their own search behavior problems so that it can become a social thinker and solve social problems. Learners are shown the bad behavior or unreasonable (unthinkables) so be of good behavior or reasonable (thinkables) through the appearance of figures depicting child unthinkables attitude. Superflex® is the human figure superhero owned in every child's mind. The goal is to assist students in self-regulating, social thinking and social skills. In addition to learning Superflex® regular teachers / general can help learners explore the capabilities of social thought.

The main premise of this curriculum is that everyone can follow to become a superhero in a situation to overcome Unthinkables Team (Baker 2011, 33). Indicators of successful implementation of the curriculum depends on the assessment by the teacher. Rate tailored to the characteristics Superflex® learning model that aims to enable students to have social skills. In the book Assessment Guide for Primary Schools (Directorate of Development Primary Schools, 2015) performance assessment is one of the skills assessment done by identifying the characteristics of learners.

The "Unthinkables" are characters that invade your brain and make your thinking inflexible and make you have unexpected behaviors (Madrigal & Winner, 2008). According to figures from the team Unthinkable Madrigal & Winner (2008) in (Rachmah, 2015) is:

1. Rock Brain: Person gets stuck on their idea of what they want to do and will not negotiate with other people. The person is not a good problem-solver and tries one

- solution that's not working over and over again. This person may be very rule bound and have rigid thinking, only seeing one way in a situation.
2. Brain Eater: Makes it hard for the person to focus on what he is doing or focus on others during interactions. The person may get easily distracted with their own thoughts or things around them.
 3. Glassman: Makes a person have a big reaction to a small problem. This person goes from 0-60 mph and quickly gets very upset often over "tiny" problems. Glass Man usually thinks things are not "fair."
 4. Mean Jean: This person is mean to other people. They insult or criticizes others. They may take things away from others, be very bossy, or try to get all the attention when others are trying to talk.
 5. D.O.F (Destroyer of Fun): This character often pops up during games or activities involving competition. The person becomes very competitive and insists on going first, playing only what he wants to play, and does not think about compromising or about how they make others feel.
 6. Space Invader: This character makes the person's body move into other people's space when others are not expecting it or do not want this. They do not realize how uncomfortable this makes others feel.
 7. Was Funny Once: This person will attempt to use a lot of humor to be funny. However, he does not realize that humor wears out pretty quickly or at times is not "funny" at all. He has trouble recognizing appropriate times for humor and may try to be funny during a discussion in a classroom or when the moment is serious and not funny or silly. Some people may get so silly, that the other children become silly also, making the group fall apart. This is called getting caught up in the "silly tornado."
 8. Energy Hare-y: This character gives the person so much energy so the person is constantly fidgeting, moving around, and doesn't think about what the people around need or how others are feeling around them. Sometimes, Energy Hare-y and Was funny once work together, which can quickly make the group fall apart.
 9. Grump Grumpaniny: Makes the person think the worst or feel like people are always unkind. He ends up believing it even when people are trying to be nice. He may also see everything as negative or bad and does not see how his emotion spreads and makes everyone feel unhappy.
 10. One-Sided Sid: This character gets the person to talk about his own set of topics or his own plan. Even when someone else brings up his interests, he just talks about his interests. He may interrupt to talk about what is on his mind, not seeing that someone may have another plan.
 11. Topic Twister Meister: This character gets the person to twist the topic around to what he wants to talk about and goes off on tangents when talking to others. This person may then go on and on about topics he wants to talk about, not realizing that others may be bored or disinterested in what he is talking about.
 12. Worry Wall: He makes the person worry or feel nervous so much about the people around him or the social situations that he or she "hits a wall" and stops being able to talk at all to the people nearby.
 13. Body Snatcher: Person wanders away from others, the group or person he is with. May also get the person to turn their body away from the group, not realizing the message they are sending to others.
 14. Un-Wonder: Stops the person from showing interest (social wondering) in others or thinking about what others may want to do. The person may not ask a lot of questions about others.

The performance assessment is an assessment that asks students to perform tasks on the real situation to demonstrate the knowledge and skills required (Directorate of Development Primary Schools, 2015). Simply put, the performance appraisal can be expressed as an assessment of the capabilities and attitudes demonstrated through an act (Wulan, 2013). Based on research conducted by Wulan (2013) the strength of performance assessment is that students can demonstrate a process that can be observed directly. Evaluation is more complete and natural for some kinds of reasoning, verbal ability, and physical skills, and be able to assess learning outcomes and complex skill. Students will be motivated to apply learning in real life situations. According to Wisconsin Education Association Council, (1996):

"Advocates of performance assessment call for assessments of the following kind: designing and carrying out experiments; writing essays which require students to rethink, to integrate, or to apply information; working with other students to accomplish tasks; demonstrating proficiency in using a piece of equipment or a technique; building models; developing, interpreting, and using maps; making collections; writing term papers, critiques, poems, or short stories; giving speeches; playing musical instruments; participating in oral examinations; developing portfolios; developing athletic skills or routines, etc"

Device performance assessment for learning model Superflex® adopt the device which has been prepared by Baker (2011) which includes basic demographic questionnaire, observation and interview guides sheet. The reason should be considered in preparing the assessment are:

(1) the criteria define for students and others the type of behavior or attributes of a product which are expected, and (2) a well-defined scoring system allows the teacher, the students, and others to evaluate a performance or product as objectively as possible. If performance criteria are well defined, another person acting independently will award a student essentially the same score. Furthermore, well-written performance criteria will allow the teacher to be consistent in scoring over time (Wisconsin Education Association Council, 1996).

Based on the results of preliminary observations made in SD Cimahi, Superflex® learning model has never been done in school. There are some students who are experiencing emotional and behavioral disorders (tunalaras) with traits include: being rebellious, irritable, undermine, disrupt and break the rules of the school. In fact, this article is part of a study entitled "Model of Learning Superflex® to Explore Social Behaviour Thinking" (Rachmah, Gunawan, & Mulyani, 2015), up to this article can be assumed as the initial research to supplement the research that has not been completed.

This study aims to (1) identify the assessment sheet for respondents who have behavior problems (2) describe the behavior of students tunalaras after getting Superflex® learning model with performance assessment refers to the curriculum Superflex® in social book thinking (Winner, 2014). The benefits gained from this research is that students can feel the effects of an impact on the performance assessments better behavior change and increase knowledge of teachers in primary schools in dealing with students who have special needs in the field of tunalaras. Other expectations, this research may add to the literature on the development of social thinking.

II. RESEARCH METHODS

The research method was the case study. In this study, researchers took one of the students named RJ with traits often teasing, irritability and a violation of the rules of the school and his friend of the same age but did not have a problem perilaku to help give an assessment to RJ. According to Creswell (2009) cases should be limited by time and the activities that take place. When the study is only one week (three times) to see the change in behavior that occurs in RJ. Instruments used in this research is interview guidelines for RJ and parents and observation sheet to see RJ behavioral change. The data obtained are analyzed in detail on the case along with the setting and then studied theory and previous research that has been conducted by Baker (2011).

III. RESEARCH FINDINGS AND DISCUSSIONS

Respondents in this study is RJ Elementary School 4th grade students who have behavioral problems in class. The training sessions conducted in private in the classroom after RJ follow the lessons. The training was conducted for 120 minutes and RJ may express his feelings at any time. The problem often faced by RJ is often angry rant and yell in class if their wishes are not met. RJ is the only child who lived in dense residential area in Cimahi. The boundary between the neighboring houses RJ is only limited by the walls of the house attached, so if the conditions and the situation monitored by the neighboring house and vice versa.

Complaints obtained from teachers is often RJ hurl harsh words. According to his parents in the neighborhood around the house allows for RJ to absorb the harsh words that every day disclosed by the people around him. RJ became grumpy because often scolded by his parents. In the category unthinkables team, RJ categorized as "Glassman" that should be developed confidence in himself as the Thinkables and become "Cool Q.Cumber" super hero figure in her that will help RJ to remain calm face everyday activities.

Other respondents are NS, NS playmates once neighbors RJ. NS is the first child of 2 (two) brothers, with a distance of 2 years old. NS has a jovial nature, adaptable and often are not offended when RJ grumpy. NS age of 10 years and is an outstanding student in his class. NS role needed to help RJ control her anger.

A. Sheets Assessment for Respondents Who Has Behavior Problems

Assessment sheet used in this study were adopted from various sources. The required assessment sheets submitted to the respondent (RJ and NS). Assessment sheets prepared are:

1. Anger Map Worksheet
(<http://www.sheffkids.co.uk/adultssite/documents/worksheets/Anger%20Map.pdf>)
2. Self Control Worksheet
(<http://empoweredbythem.blogspot.co.id/2013/04/self-control-is-big-issue-with-some-of.html?showComment=1365559669519#c573555574862066461>)
3. Observation sheet about social skill
(<http://www.escambia.k12.fl.us/pbis/fbadata/SOCIAL%20SKILLS%20CHECKLIST%20-%20ELEMENTARY.pdf>)

Anger map worksheet given to RJ as a means to express angry feelings at the time. Assisted by NS, RJ depicting her in a moment of anger, words are spoken, the behavior when angry, angry effect on the body, how to deal with anger and what to do to others when he was angry. Just like Anger Map, Self-Control worksheet given to RJ. RJ asked to do things that are requested in paper work when he felt angry assisted by NS.

Researchers create a worksheet to identify the anger that arises. RJ asked to check the mark on every statement that makes him angry. Once filled, the researchers provide an overview of how to overcome the anger that emerged in accordance with perceived by RJ. According to Madrigal & Winner (2008) How to cope with anger can be overcome with:

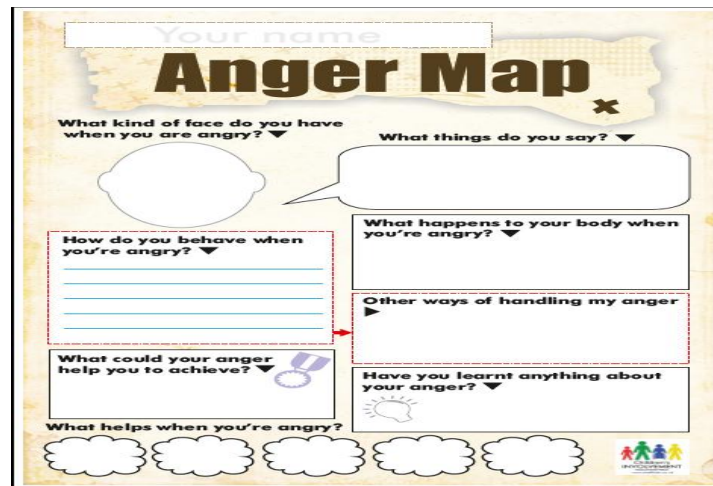
1. STOP and think about the problem
2. Tell yourself to take a break and walk away
3. Ask for help
4. Take a deep breath to calm down

Observation sheet filled in by the researchers to see changes that occur in RJ every meeting. By filling out the observation sheet, researchers know the changes that occur at the next meeting.

B. Student Behavior After Getting Superflex® Learning Model with Assessment of Performance

At the first meeting dated 25 April 2016, researchers looked at the behavior of RJ upon learning outside the classroom for two hours. RJ seen often angry at small things like you can not do the task, did not find an eraser, knock friends and others. At the time reprimanded by the teacher, RJ yell and argue that a given task is too difficult, eraser taken friends, and always said that everything that happened was not her fault.

RJ study hours starting from 07.00 until 12:00. After resting for 1 hour, accompanied by parents and NS, RJ talk relaxed and given paperwork Anger Map. Working Paper Folders given Anger can be seen in the following figure:



The worksheet is titled "Anger Map" with a space for "Your name" at the top. It contains several sections for reflection on anger:

- What kind of face do you have when you are angry?** (Accompanied by a blank face outline)
- What things do you say?** (Accompanied by a speech bubble)
- How do you behave when you're angry?** (Accompanied by a list of horizontal lines for writing)
- What happens to your body when you're angry?** (Accompanied by a list of horizontal lines for writing)
- Other ways of handling my anger** (Accompanied by a list of horizontal lines for writing)
- What could your anger help you to achieve?** (Accompanied by a list of horizontal lines for writing)
- Have you learnt anything about your anger?** (Accompanied by a list of horizontal lines for writing)
- What helps when you're angry?** (Accompanied by five cloud shapes for writing)

At the bottom right, there is a small logo for "ADDITIONAL RESOURCES" and a website address.

Figure 1. Anger Map Worksheet

(<http://www.sheffkids.co.uk/adultssite/documents/worksheets/Anger%20Map.pdf>)

Judging from the results of the charging is done RJ, on the question "Have you learned anything about your anger? RJ answered with "I do not know" and did not know how to cope with anger. RJ said, "ugly face", "want to cry", "want to throw" and "wanted to scream" when he felt angry. RJ wants teachers and friends in the classroom, according to him and his wishes fulfilled.

After identifying the anger owned RJ, researchers deliberately provoke anger RJ by taking paper and scribble paper. RJ looks angry, and immediately slammed the pen is being held. Then the researchers asked RJ to breathe deeply several times and walk quietly. RJ looks a bit quiet though his face looks still angry. RJ researchers asked about feelings when asked to take a deep breath repeatedly, but RJ did not want to answer and showed an angry countenance. Researchers suggested that if feels to be angry, try to take a deep breath over and over. From the first meeting, researchers can identify ways that can be done in overcoming anger.

The second and third meetings held respectively on 27 and 29 April 2016, with the same observation. Although there has been no significant change during the process of learning in the classroom, but teachers always remind RJ to breathe deeply whenever RJ would get angry and yell, although still RJ angry and shouting.

Outside the classroom, researchers asked RJ superhero whichever is preferred. RJ replied that he liked the Hulk, the man who turns into a green creature and stronger by the time he was angry but defenders of truth. At the second meeting of the investigators told me that RJ can be Kool Q.Cumber figure who could help cool the anger and powerful as Hulk. At a meeting the third, researchers featuring Hulk movie and asked RJ to express their opinions about the films they watch.

Observations on the second and third meetings show that RJ showed little change in dealing with anger. Initially RJ can be angry, throwing objects in his hand. But at the third meeting of missiles has been reduced only anger and the desire to shout still can not be controlled. Limitations of time, due to meet the deadline of writing this paper makes researchers pause, handing a note to the teacher to try to do things that are done by researchers.

The results of these studies indicate that anger in students can be addressed with appropriate measures. Superflex® strategies suggests that anger can be overcome by:

1. Identify the size of the problem (1-10) and what would be an expected reaction to match the size of the problem.
2. Self-Talk: "I am starting to get mad. I need to move away and take a break." "This not a big deal." (Madrigal & Winner, 2008)

Research conducted not representative enough for a short time. Supposed to invade behavior, should be done in a long time and do not rush. This will affect the results of research and break the rules in learning Superflex®. In doing Superflex® learning there is 10 (ten) things that can and can not do, it can be seen in the following table:

Table 1. Do and Don't in Superflex®

No	Do This	Not This
1	DO start by teaching how to be a social OBSERVER (Social Detective) before introducing them to their Superhero (Superflex).	DON'T rush the concepts.
2	DO emphasize that Superflex isn't a comic book character that lives "outside" us all - SUPERFLEX IS YOU!	DON'T use Superflex if students don't have a clear understanding of the difference between fantasy/pretend and reality.
3	DO empower kids to figure out their own team of Unthinkables and Thinkables. You'll have a greater level of buy-in if kids feel like they are figuring out their own challenges and strengths (self-awareness) and then a plan to overcome (behavior change).	DON'T use UNTHINKABLE terms to describe the child. Avoid saying things like, "You're being a Glassman" or "Stop being such a Rock Brain."
4	DO link Superflex and the Unthinkables and Thinkables to academics and content standards.	DON'T teach about the Unthinkable characters without teaching strategies to defeat them!
5	DO use the Social Detective and Superflex strategies with the general education population too!	DON'T use the Superflex curriculum (as is) with students with limited verbal abilities or those who are just emerging into language (single words or short phrases).
6	DO be cautious of what outcomes you expect when using Superflex with younger students.	DON'T use Superflex concepts and strategies with preschoolers.
7	DO link Superflex concepts to Social Emotional Learning (SEL), Positive Behavioral Interventions and Supports (PBIS), and Response to Intervention (RTI).	DON'T turn Superflex concepts into a behavior plan or teach the concepts in a purely behavioral way.
8	DO take plenty of time to teach from many different angles.	DON'T rely on the Superflex handouts, the Superflex poster, and the Rock Brain comic book as your only tools to teach and build Superflex concepts.
9	DO involve parents/caregivers and make sure all know the vocabulary and are clear about expectations. Behavior change is often SLOW.	DON'T let kids run away with the vocabulary in a way that reflects violence and STOP using if kids become obsessed with the negative behaviors of the Unthinkables.
10	DO have fun and be creative!	DON'T sell what you create!

Source: (Crooke & Winner, 2015, hal. 3)

Implementation of the assessment to students will go well if the teacher can collaborate with other teachers and principals. Because conceptually one of the principal tasks adal guide and motivate teachers to engage in activities that support the development of students. So that the principal is able to assist teachers in developing their professional ability (Fazilla & Marisa, 2015)

IV. CONCLUSION AND SUGGESTIONS

A. Conclusion

1. The performance assessment can be done in learning Superflex®. Instruments tailored to the needs assessment (who, how and what). In this study, the research instruments taken from various sources that are downloaded from a trusted source and consistent discuss about social thinking, Superflex® and social skills.
2. There is little change in the behavior of students after getting treated in accordance with the agreements that have been made. Changes respondents felt at the time was no longer throwing objects at some moment of anger that do.

B. Suggestions

1. Limitations conduct research on the use of assessment performance make research results have not been in line with expectations. For that to the next researcher to redevelop the results to the long term and following the steps that are required in theory.
2. Teachers can be more active and personal in approach with students who have problems with anger. Most students feel cared for without feeling being studied.

REFERENCES

- [1] Baker, K. E. (2011). Detectives and Superheroes: A Pilot Study Teaching Flexible Thinking in Social Situations to a Child with High Functioning Autism. *Thesis*. St. Catharines, Ontario: Faculty of Social Sciences, Brock University.
- [2] Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approach*. Thousand Oaks California: Sage Publication.
- [3] Crooke, P., & Winner, M. G. (2015, Maret). *10 DOs and DON'T for Teaching Superflex*. Think Social Publishing,inc.
- [4] Directorate of Development Primary Schools. (2015). *Assessment Guide for Primary Schools*. Jakarta: Directorate of Development Primary Schools, Ministry of Education and Culture.
- [5] Fazilla, S., & Marisa, R. (2015). Teacher Development Capabilities in Doing Assessment Portfolio Primary School Students. *Jurnal Penelitian Pendidikan (Educational Research Journal)*, 15(3), 11-22.
- [6] Joyce, B., Weil, M., & Calhoun, E. (2009). *Models of Teaching* (8 ed.). (A. Fawaid, & A. Mirza, Trans.) Yogyakarta: Pustaka Pelajar.
- [7] Madrigal, S., & Winner, M. G. (2008). *Superflex: A Superhero Social Thinking Curriculum*. www.socialthinking.com.
- [8] Rachmah, H. (2013). Model of Integrative Learning in Indonesia: an Alternative. *Seminar Nasional Pendidikan Sejarah "Menyongsong Kurikulum Sejarah 2013"*. Jakarta: APPS dan Jurusan Sejarah FIS UNJ.
- [9] Rachmah, H. (2015). Superflex Learning Model to Improve Social Skills. *Paper for Guangzhu University*. Unpublish.
- [10] Rachmah, H., Gunawan, R., & Mulyani, M. (2015). Model of Learning Superflex to Explore Social Thinking Behaviour. *Competitive Research Grant Proposals*. Cimahi: STKIP Pasundan.
- [11] Winner, M. G. (2014). *Superflex, the Team of Unthinkables and the Five-Step Power Plan*. Retrieved Maret 28, 2015, from social thinking website: <http://www.socialthinking.com/what-is-social-thinking/superflex-the-unthinkables-and-the-five-step-power-plan>
- [12] Wisconsin Education Association Council. (1996). *Performance Assessment*. Retrieved from www.learner.org: <https://www.learner.org/workshops/socialstudies/pdf/session7/7.PerformanceAssessment.pdf>
- [13] Wulan, A. R. (2013). Penilaian Kinerja dan Portofolio pada Pembelajaran Biologi. *Artikel Pendidikan Biologi*.

THE IDENTIFICATION OF TEACHERS DIFFICULTIES IN IMPLEMENTING OF 2013 CURRICULUM AT ELEMENTARY SCHOOLS

Ika Maryani¹⁾ Sri Tuter Martaningsih¹⁾

¹⁾ Elementary Education Department, Ahmad Dahlan University
Email: ika_mail87@ymail.com

Abstract - This research is based on the teachers difficulties in implementing the 2013 Curriculum at Elementary Schools. The purpose of this research is to describe the obstacles faced by the teachers in implementing of 2013 Curriculum in elementary school. This research used descriptive method. The results of this research indicate that the majority of elementary school teachers still have a low understanding of authentic assessment concept. Techniques and instruments that exist in the teacher's guide are less appropriate to the school culture. Teachers are still difficulties in preparing self-assessment and peer assessment instruments. They find difficulties in measuring and describing the attitudes and learning performances.

Keywords: *assessment, 2013 Curriculum, elementary School.*

I. INTRODUCTION

Law No. 14 of 2005 on Teachers and Lecturers defines that teachers are professional educators with the primary task of educating, teaching, guiding, directing, training, assessing, and evaluating students on early childhood education, formal education, primary education, and secondary education [5]. Teachers are not only responsible for educating and facilitating the learning process but also have to adapt to the various changes, especially changes in the curriculum.

Curriculum changes should ideally be done continuously in accordance to the challenges that students will face in the future. Curriculum change is also a result of the development of society. We do not want to build a separate generation to the development of society. We are educating the next generation that will live in a different era with us. We mature them through education which is not worn out, which the charge contained in the curriculum [4].

2013 Curriculum is the development and refinement of the previous curriculum to respond to various internal and external challenges. One of the reasons for the importance of 2013 Curriculum is that young generation needs to be prepared in the competency of Indonesia attitudes, skills, and knowledge. The scientific approach in 2013 Curriculum is an attempt to meet these needs. Implementation of 2013 Curriculum is seen as a much needed. Therefore, the curriculum has been implemented on a limited basis at the beginning of the school year 2013-2014. It is implemented in the schools that meet the requirements, and it's determined selectively. The target of a year trial period of 2013 Curriculum and the plan to implement it fully the next year was experiencing obstacle and policy changes. Some schools go back to the School-Based Curriculum but most are still implementing 2013 Curriculum. It is in accordance to the readiness of each school.

Based on data obtained from the Electronic Systems Monitoring Implementation of 2013 Curriculum, there are 2,865 elementary schools in Indonesia that become the target school to implementing 2013 Curriculum, while in Yogyakarta there are 64 elementary schools starting the implementation of 2013 Curriculum in the academic year 2013/ 2014 ago. Based on the results of preliminary observation conducted by researchers at the end of December 2013, many teachers in the school that implementing 2013 Curriculum, complain about the difficulty in implementing the system and rules specified in the 2013 Curriculum.

The readiness of the implementation of Curriculum 2013 can be viewed on the commitment of teachers, teachers' understanding, and support of learning resources [6]. Teachers'

understanding of the essence of Curriculum 2013 is an important factor, so that Curriculum 2013 can be accomplished sustainable and systematically. Teacher's understanding influences the implementation of 2013 Curriculum's readiness. If the teachers understand the rules that apply in Curriculum 2013, then all questions and problems in learning will be missed. On the contrary, if the teacher does not know how the essence of Curriculum 2013, the implementation would have many difficulties.

Based on the explanation above, research on the teachers' problems identification in implementing the 2013 Curriculum needs to be done to determine the overview of teachers' understanding of the 2013 Curriculum. The results of this study will be input for Teacher Training Institution to take remedial action in overcoming the difficulties and problems faced by the teachers in the field.

II. RESEARCH METHOD

This research is qualitative using descriptive method. Through this research, the overview of the implementation of 2013 Curriculum can be figured out. Next, the factors that cause the certain phenomena can be analyzed. This method is an ex post facto research. The sampling technique in this study using purposive sampling technique. The samples are Muhammadiyah elementary schools from Sleman, Yogyakarta Special Province. To collect the data, the instrument is a depth-interview and self-assessment technique. The collected data is then processed, reduced, analyzed to get a conclusion.

III. RESULTS AND ANALYSIS

In this research, several variables are used to indicate the level of teachers' understanding on the 2013 Curriculum, there are the understanding of scientific approach, authentic assessment, and thematic integrative learning. Table 1 below shows the description of the teacher's perception on 2013 Curriculum.

Table 1. Perception Category of Teachers to 2013 Curriculum.

Variable	Mean	Category
Understanding on the scientific approach	65.26	Low
Understanding on the authentic assessment	67.18	Low
Integrative understanding on the thematic learning	75.62	Average
Mean	69.35	Low

The research results of the elementary school teachers' perception in Sleman, shows that teachers' understanding on the 2013 Curriculum is still low, with mean scores 69, 35.

When observed from each variable, the understanding of scientific approach is consisting of a scientific approach concept and its implementation in learning. While understanding the authentic assessment consists of authentic assessment concept, authentic assessment principles, authentic assessment characteristics, authentic assessment techniques, and instruments preparation in an authentic assessment. Understanding the integrated thematic learning consists of thematic learning concepts, preparation of thematic learning plan, and thematic learning implementation. Descriptions of each variable are shown in the following explanation.

A. Understanding Scientific Approach.

Based on the questionnaire analysis, teachers' understanding of the scientific approach, showed a tendency of low perception toward scientific approach. This condition is indicated by a mean of 65.26, it means in the low category. Teachers' lack of understanding of scientific approach could not be separated from the teachers' inability in scientific approach concept and its implementation in learning. This data was measured using questionnaire toward the understanding scientific concepts, teachers' Lesson Plan assessment, and observations to the scientific learning process carried out by teachers.

Data of teachers' understanding categorization on scientific approach concept is shown in figure 1.

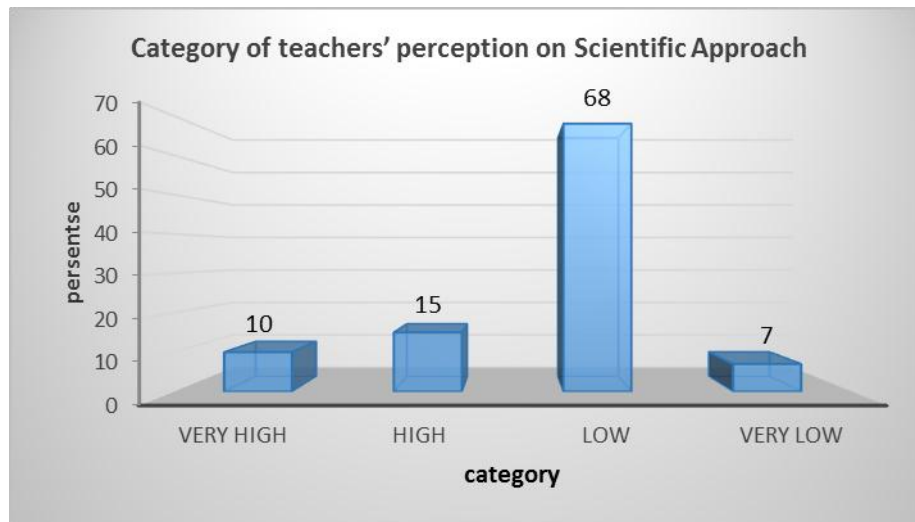


Figure 1. Category of teachers' perception on Scientific Approach

Figure 1 explains the overview of teachers' perceptions of a scientific approach. Teachers' perception which are at very high category are 10 teachers (10%), high category are 15 teachers (15%), low category are 68 teachers (68%), and very low category are seven teachers (7%). Based on the mean score obtained from a sample of 100 elementary school teachers in the district of Sleman, it could be concluded that the teachers' understanding of the scientific approach is at the low category.

Based on observations and depth interviews to 40 elementary school teachers in the district of Sleman, obtained information that the teachers' difficulties in developing the scientific approach are derived from the following aspects:

1. Teachers used to implement teacher-centered learning while the scientific approach has the characteristics of student-centered learning. It needs necessary refraction and perceptions change between teachers, students, and parents about the approach. Syntax scientific approach which consists of observing-questioning-associating-experimenting-networking becomes difficult because students have not been accustomed to do it in learning. In scientific approach, students are expected to find the environment concept, while not all schools have an environment that can be used as a learning resource. The teacher is required to entirely understand the characteristic of scientific approach and the purpose of each phase as it is expected by the 2013 Curriculum as it has already been disseminated through workshop and training. In addition, in order to cope with the time management that might be the concern of elementary teacher during teaching process, collaborating with peers can be a solution to improve the effectiveness of the teacher's teaching practice. Partners can evaluate and give some feedbacks of the teaching practice arranged by the teacher. Therefore, it strengthens the sense of reflective teaching for teacher so that she or he is able to find the strength and weakness of their teaching practice. Creativity is also vital for the teacher in conducting valuable and pleasant teaching practice [7].
2. Teachers opine that the scientific approach takes longer time, requires more time in preparation, students' assessment becomes more complicated, the children underachieving will have difficulty in learning, scientific approach less suitable for the difficult lesson material, students feel their task (homework) are too much, and it takes time to change student' habits to have scientific attitude, critical thinking and creative intelligence. Syntetics technology is a systematic way of designing, carrying out and evaluating the teaching learning process. A Syntetics technique makes scientific attitude, critical thinking and creative intelligence more effective, understandable and meaningful [2].
3. Teachers opine that to implementing scientific approach, there are contributing factors such as: the availability of a reference work, the availability of educational props, the availability of learning facilities, teachers' capabilities and professionalism, availability of accessing information both for the teachers and for the students, the teachers training, utilizing ICT, students' readiness to learn, and school readiness. Conditions in the field indicate that not all schools and teachers are able to provide these factors. Limitations of supporting facilities at a particular school become an obstacle to the implementation of the scientific approach.

4. Teachers opine that they have trouble applying the scientific approach as follows: not all of the lesson material is easy to be done by the scientific approach, difficulty in apperception, difficulty in assessment based on scientific approach, difficulties in steps/process of questioning, and did not understand the ins and outs of the foundation and philosophy of scientific approach.

B. Understanding authentic assessment

The results of the data obtained through a closed questionnaire containing 30 items of questions, the maximum score for each item is 4 and the lowest score is 1. Based on the mean score obtained from a sample of 100 elementary school teachers in the district of Sleman, it could be concluded that the teachers' understanding of the authentic assessment is in a low category. The tendencies are seen from the position of an objective mean 67.8 which are in the low category.

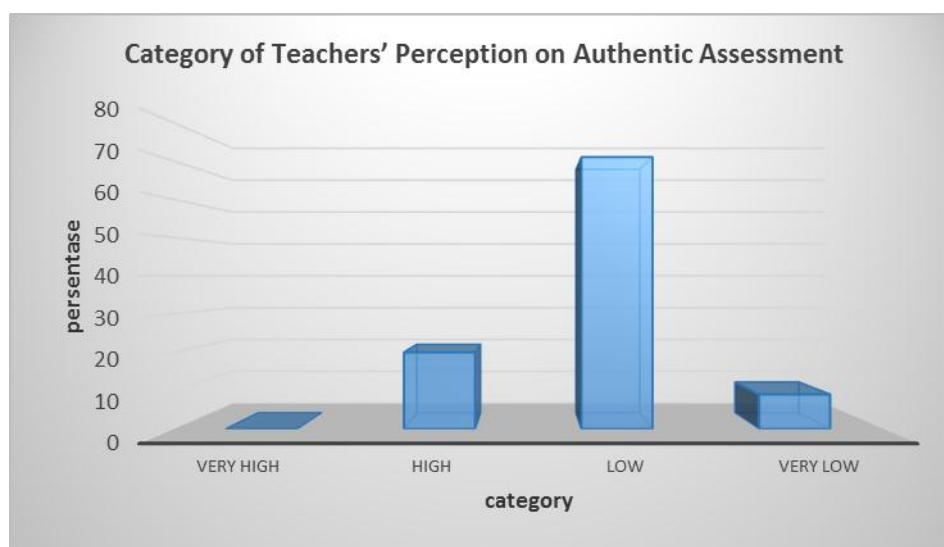


Figure 2. Category of Teachers' Perception on Authentic Assessment

Figure 2 explains the overview of teachers' perceptions on authentic assessment if used in ideal criteria. Teachers' perception which are at very high category is 0 teacher (0%), high category are 20 teachers (20%), low category are 71 teachers (71%), and very low category are 9 teachers (9%). Based on the mean score obtained from a sample of 100 elementary school teachers in the district of Sleman, it could be concluded that the teachers' understanding of the authentic assessment is in a low category.

Teachers' low understanding on authentic assessment related to the teachers' understanding of the evaluation concept. Most respondents give the same sense that the evaluation considered as a grading. The learning results as if still knowledge oriented without regard to other domains that influence it. This notion is not entirely wrong. Witherington in Zainal Arifin (2011) said: "*an evaluation is a declaration that something has or does not have value*". Evaluation is defined as a grading [1]. However holistically, the evaluation should demonstrate a continuous and systematic process and covers all areas such as knowledge, skills, and attitudes. The implication of this understanding is that the teachers must be skilled in assessment planning, implementing, managing, and compiling learning outcomes report.

On the authentic assessment concept, assessment focused on what learners could do to solve real-world problems, not just the school world. This type of assessment uses a variety of holistic ways to reflect the attitudes, knowledge, and skills. Attitude assessment on 2013 Curriculum includes assessing the spiritual and social attitude. Spiritual attitude is the attitude toward God, which of course contains assessments in terms of worship/praying. Social attitude is the attitude toward people, containing attitudes in social interaction. Attitude assessment could be done by observation, self-assessment, peer assessment, teacher's notes, and journals. Based on the questionnaires analysis results conducted by the researcher, most of the respondents felt the lack of understanding on the implementation of this attitude assessment. Limitations of the teachers' observation become a major factor that attitude assessment is less maximal. For schools with large classes (> 20 students), the observation becomes very difficult.

As a result, teachers do not use this technique or even may just do the scanning of the students who stand alone so that assessment becomes subjective. For the self-assessment techniques, journals, as well as peer assessment, teachers are still found difficulties in the instruments development, the assessment implementation, as well as how to analyze the assessment results.

Domain knowledge assessment was conducted to determine the achievement of learning purpose in the cognitive domain. In this assessment, teachers are accustomed to using written and oral tests techniques. The difficulty faced by teachers in the knowledge assessment is in developing core competencies, learning outcomes, and indicators. Teachers are still found difficulties to develop thematic question because they used to be too fixated on the subjects. Therefore, in practice, the knowledge domain assessment is subjects' assessment.

Beside attitude and knowledge assessment, teachers also had difficulty in assessing the skill domain. Every student should be observed one by one according to the skills indicators. However, the researcher limited ability makes observations of this technique becomes less than the maximum. Teachers' ability in developing assessment instruments also becomes an obstacle. Teachers generally found it difficult to develop the skills indicators in which observed. In addition, the teachers' ability to analyze and interpret the results of the assessment in the form of a descriptive report is still very low.

An understanding of authentic assessment system that previously understood as a form of classroom assessment, in general, still needs to be increased, including the implementation from the planning to the follow-up. Teachers used to evaluate learning outcomes using the written test in the form of daily tests, the middle semester test, and the final test. The teachers still not understand the assessment types that can be done to measure student competencies. Various assessment methods are offered in the 2013 Curriculum became something novelty for teachers.

An understanding of the various types of evaluation in authentic assessment is relevant to the competence to be assessed. Characteristics of the object, the target, the appropriate evaluation techniques, and the development of organizing each type evaluation still need to be improved. Performance assessment, project assessment, portfolio assessment, and written tests assessment is authentic assessment types that can be used to assess cognitive abilities and skills. However, performance assessment, project assessment, and portfolio assessment has not been understood by the teachers, so that they cannot implement it. Most teachers understand the written test assessment, however, the diversity of types, the ability to design the instrument, instrument quality testing, implementation, processing, up to analysis still requires further training and guidance. Similarly, in terms of mastery of performance assessment system, portfolio assessment, project assessment and etc, teachers need to increase their knowledge as well as practical implementation

Teachers' skills in developing and organizing authentic evaluation require training and assistance in a specific period of time, so evaluation process could be fully understood and the teachers could implement it. The formulation of evaluation instruments and its use in measuring students' competency require teachers 'creativity.

In the stage of evaluation system planning, observations were made on the teachers' ability to develop indicators moreover indicator of the ability to think critically, they need to practice in order to implement it. The evaluation system planning as a whole, which should be reflected in the Lesson Plan also not yet apparent. Data evaluation draft also needs to be developed especially when there is a change in the report assessment system in the implementation of 2013 Curriculum.

The teacher also opines that Spiritual and social competence assessment still encountered obstacles. The teachers do not understand the assessment mechanisms of both basic competencies. During the assessment implementation phase, teachers often unprepared, although teachers are familiar toward their students, the objectivity and accuracy of the assessment need to be better.

In the processing and follow-up, teachers should develop students' abilities in a narrative description. The ability to narrate assessment demand good ability and it turns out that the teachers had difficulties in this regard, given they usually more oriented in giving numbers [3].

C. Understanding the integrative thematic learning

Based on the analysis of the teachers' understanding on thematic integrative learning, shows that teachers quite understand thematic integrative learning. This condition is indicated by a mean of 75.62 or in the average category.

Categorization data of teachers' understanding on the integrated thematic learning concept is shown in figure 3.

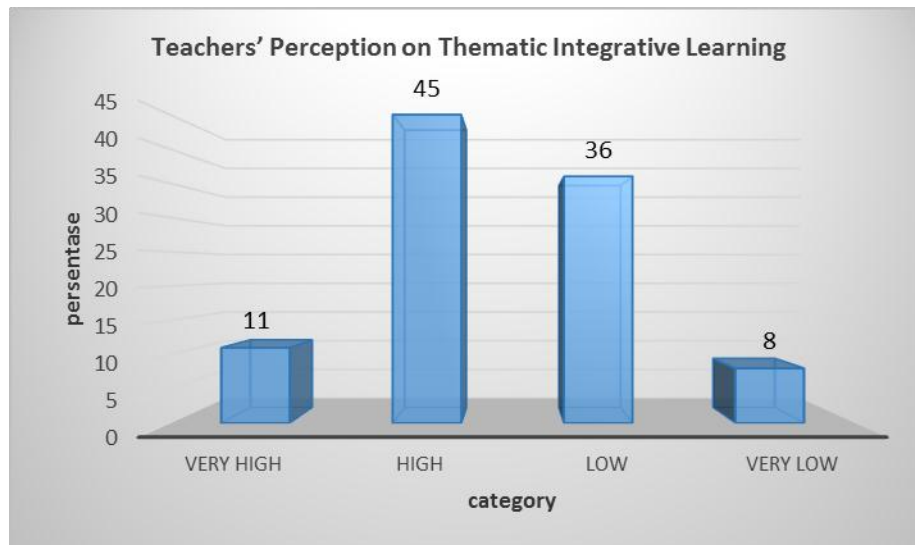


Figure 3. Teachers' Perception on Thematic Integrative Learning

Figure 3 explains the overview of teachers' perceptions on the thematic integrative learning. Teachers' perception who are at very high category are 11 teachers (11%), high category are 45 teachers (45%), low category are 36 teachers (36%), and very low category are 8 teachers (8%). Based on the mean score obtained from sample of 100 elementary school teachers in the district of Sleman, it could be concluded that the teachers' understanding on the thematic integrative learning is in the average category.

Based on the results of depth interviews with 40 respondents, it provided information about the constraints faced by teachers in the integrative thematic learning, there are:

1. The time is limited. For example, the time allocated for a subject is one day. While sometimes in one learning, the lesson materials are so much that is not possible to be learned in one day. In addition, if students still do not understand about the lesson materials that have been studied, the teacher cannot force students to continue to learn the next lesson material, because the teachers want students to really - really understand the lesson material that being studied. As for the assessment, teachers also experiencing difficulties because there are many aspects need to be assessed.
2. In planning thematic learning, teachers found many obstacles in developing indicators that in accordance to the learning outcomes; formulate learning purpose to meet the elements of Audience, Behavior, Condition, and Degree (ABCD); make the assessment format on the Lesson Plan; create interesting learning activities or scenarios; and developing learning materials that contained in the syllabus and teacher books.
3. On the learning implementation using a thematic integrative approach, teacher found difficulty in growing students' logic; courage to communicating the results of the work or task that has been done; sharpening students' memory; mastery of the class, performing activity steps; lack of assessment time; and developing sources and media in the classroom.

IV. CONCLUSION

Based on the research, we find that the problems of elementary school teachers in implementing the 2013 Curriculum caused by inability of the teachers to understand the concept of scientific approach, authentic assessment, and thematic integrative learning. There are many factors that cause these problems, i.e the teachers have not been accustomed to student-centered learning, the limitation of learning facilities, and the limited ability of teachers to the

component of 2013 Curriculum. Therefore, we recommend that the training and workshops followed by intensive mentoring for elementary school teachers on the implementation of the 2013 Curriculum.

ACKNOWLEDGMENT

This research was funded by Research and Development Bureau of Ahmad Dahlan University.

REFERENCES

- [1] Arifin, Zainal. "*Evaluasi Pembelajaran Prinsip, Teknik, Prosedur*". Bandung: PT. Remaja Rosdakarya". 2011.
- [2] Candrasekaran. "Developing Scientific Attitude, Critical Thinking and Creative Intelligence of Higher Secondary School Biology Students by Applying Synectics Techniques." *International Journal of Humanities and Social Science Invention* 3 (6): 01-08, 2014.
- [3] Maryani, Ika, and Sri Tuter Martaningsih. "*Persepsi Guru Sekolah Dasar terhadap Sistem Penilaian pada Kurikulum 2013*". Research Report, Yogyakarta: UAD. 2014.
- [4] Sukmadinata, Nana Syaodih. "*Pengembangan Kurikulum Teori dan Praktek*". Jakarta: Rosdakarya. 2008.
- [5] Undang- undang No. 14 Tahun 2005. n.d. "*Guru dan Dosen*". Jakarta: Pustaka Pelajar Offset. 2005.
- [6] Untari, Rahmania Sri, Amat Mukhadis, and Waras. "kesiapan Guru SMK Program Keahlian Teknik Informatika dalam Pelaksanaan Kurikulum 2013." *Teknologi dan Kejuruan* 38 (1): 1-14. 2015.
- [7] Wahyudin, Achmad Yudi, and Didi Sukyadi. "A Closer Look at the Implementation of the Curriculum 2013 in Indonesia: Should the Scientific Approach Be Used in EFL Classroom?" *RJES* 2 (2): 56-70. 2015.

AEROBIC GYMNASTICS, FITNESS, AND ACADEMIC GRADE OF HEALTH DIPLOMA STUDENTS FROM REMOTE AREAS IN INDONESIA

Lucky Herawati¹, Maryana², Suharyono³

¹⁻³ The Polytechnic of Health of Yogyakarta, Tatabumi Street no: 3, Banyuraden, Gamping,
Sleman, Indonesia, post code 55293
email: dok031204@yahoo.co.id

Abstract. The study was aimed to identify the influence of aerobic gymnastics toward the fitness and academic grade of students. A quasi experiment with pre-test post-test with control group design was conducted. As the study sample were 79 students from Health Diploma 3 Study Program of 2013 enrollment year, who are originated from remote regency in North Kalimantan Province. They comprised of 39 students from "KG" Study Program as the treatment group (29 males and 10 females), and 40 students from "K" Study Program as the control group (26 males and 14 females). Aerobic gymnastics as the treatment was held in group every morning, guided by an instructor, five days a week in duration of four weeks on July 2014. The observed variables were students' fitness and academic grade. The fitness was measured through transferring the time needed by the students to reach a distance of 1.6 kilometer into maximum oxygen volume values; meanwhile the academic grade was the grades obtained by the students from Human Basis Need subject, i.e. mid-term II as the pre-test value and final-term II as the post-test value. Those grades were retrieved from documents of academic administration. The data were analyzed by using Paired T-test and Multivariate Analysis of Covariance test. The results showed that the aerobic gymnastics which were carried-out by the students significantly affect their fitness and academic grades after controlling the factors of age, gender and parental education.

Keywords: *academic grade, aerobic gymnastics, fitness, health diploma students*

I. INTRODUCTION

In 2009-2011 Polytechnic of health of Yogyakarta received students Diploma 3 study program under cooperation with some remote, border and archipelago areas i.e. one regency in West Kalimantan Province, one regency in West Papua Province, and one regency in North Kalimantan Province. In general, the academic performances of those students were below the class average. Based on the graduation data, one of Diploma 3 study program showed low percentage of graduate who achieved Grade Point Average (GPA) more than or same with 2.75. In 2009, 2010, and 2011, GPAs of the graduate of the cooperation class with one of the remote, border and archipelago area in West Kalimantan province were 15.79 percent out of 38 students, 13.51 percent out of 37 students, 28.26 percent out of 46 students, and meanwhile the GPAs of the class from West Papua area was 0 percent out of 11 students.

In 2013, Polytechnic of health of Yogyakarta, received 105 new students from the remote, border, and archipelago area in North Kalimantan Province, distributed in 5 different diploma study programs. The result of learning evaluation for semester I in 2013 at those study programs showed that the academic performance of the students were of concerned. Based on data of learning evaluation semester I from diploma 3 study program "KG" indicated that no one (0 percent) passed two subjects in 2013. Remedial effort alone couldn't repaired the condition, as reflected from what had been done before to two cooperation classes from remote area in West Kalimantan and West Papua Provinces. If this condition is not attentively managed, it will continue to produce the low Grade Point Average of the graduates and the low percentage of

timely study period. This situation will also affect to the accreditation institution score. Another impact that will arise is extra budget, should be allocated by the local government that fund them.

This cooperation class is one of the priority programs of the Centre of Health Human Resources Empowerment of Ministry of Health for providing the health personnel in the remote, border, and archipelago areas.

The problem faced The Polytechnic of health of Yogyakarta was how to upgrade the cognitive ability of the cooperation class students of remote, border and archipelago area?

Since 2000, researchers in neuroscience were interested in developing studies about physical activity (personal body fitness, sport) correlated with cognitive ability^(1, 2, 3, 4). Most of the previous studies showed that both factors are correlated. However, previous findings had not yet explored students with limited cognitive ability as research subjects. Some researchers claimed that increasing brain function, physical fitness is required^(5, 6).

In term of sport and fitness, in 2011 Health agency of Yogyakarta Province had socialized aerobics through the leaflet "Kartu menuju sehat", produced by National Health Department (7). The aim of the research was to determine the effect of student's participation in aerobic gymnastic on their physical fitness and academic grade.

II. METHODS

The study design was quasi experiment with pretest posttest with control group. The study population was 105 students of cooperation class of remote, border, and archipelago area in North Kalimantan Province. The study sample was 79 students, selected by cluster random sampling technique, consist of 39 students from diploma 3 "KG" study program as the experiment group and 40 students from diploma 3 "K" study program as the control group. They lived in dormitories provided by each study programs.

Independent variable was aerobic gymnastic, namely aerobic exercise type 2⁽⁷⁾, carried out every morning with duration 40 minutes includes the warming-up and cooling-down, led by an instructor. The aerobic gymnastics were conducted 5 times a week in a month i.e. between mid-test and final-test period. Participation of students in aerobic gymnastic was monitored by researchers using attendance list.

Dependent variables were physical fitness and academic grade of the students. Physical fitness is fit condition, measured by recording the travel time gained by a respondent when doing briskly walk at constant speed to reach 1.6 kilometer, using a stop watch. The gained travel time of the respondents then were transferred into maximum oxygen volume value^(7, 8). The physical fitness measurement was measured twice, i.e. pretest and posttest for the two groups.

The academic grade is one of the indicators of thinking capacity. In this study, the academic grade was the grade of the human basic need subject which was attained by each respondent, and comprise of mid test score as the pretest and final-test score as the posttest, obtained from academic administration affairs as the secondary data. The human basis need subject was chosen as the subject in this study because students of the two groups followed this subject with the same credit weight i.e. 3 credits in semester II.

There were 3 variables assumed as the confounder, namely age, sex, and parental education⁽⁹⁾. Age is a number in year unit, counted from birth until the last birthday, measured by using questionnaire and self-reported. Sex is physical differentiation for male and female, measured by questionnaire and self-reported. Parental (father and mother) education is presented by duration of formal education that had been attended until they got a certificate and expressed by a number in years. Parental education scoring was: score of not graduating from elementary is 0, score of elementary school is 6, score of junior high school is 9, score of senior high school is 12, score of academy is 15, score of Bachelor is 16, score of master 18, and score of doctor is 22.

The data was analyzed by using Paired T test and Multivariate Analysis of Covariance test as Telles et al (2013) did, at 95 percent level of confidence; after knowing that the data was normally distributed (p value more than or same with 0,05), with controlling age, gender, and parental education.

This study was approved by Ethics Committee of Medical Faculty of University of Muhammadiyah of Yogyakarta in 2014

III. RESULTS

Characteristics of subject in the two groups were men, 20 years old, and 9 years for the formal education of their parent or similar with junior high school. Test of equality of the characteristics was independent T-test at 95 percent level of confidence. The result showed that no different characteristics between the two groups (see Table 1).

Table 1. Characteristics of Respondens in The Experiment and Control Groups

Groups	Gender (M;F)	Mean \pm SD		
		Age	Father's education score	Mather's education score
Experiment (n=39)	29; 10	19.79 \pm 1.73	9.33 \pm 3.71	9.56 \pm 3.58
Control (n=40)	26; 14	19.82 \pm 1.51	9.80 \pm 3.19	8.15 \pm 3.31
<i>p value</i> ^a		0.469*	0.194*	0.155*

^a Independent t-test

* Not significant defferent

A. Aerobic Gymnastic

The students of the experiment group done the aerobic gymnastic every morning at 5.30 A.M, 40 minutes, include worming up-cooling down, five times in a week, for a month in July 2014. The aerobic gymnastic was held on the yard in front of the campus "KG", located at Kyai Mojo Pingit street, 56 Yogyakarta. All students followed the aerobic gymnastic from the beginning to the end of the study, as they had been signed on the informed consent.

B. Physical Fitness

Physical fitness was expressed by maximum oxygen volume with measurement unit is milliliter per kilogram body weight per minute. The fitness of the students in the experiment group indicated improvement after they followed the gymnastic aerobic, i.e. 3.4 milliliter per kilogram body weight per minute or 14.76 percent. In contrast, the fitness of the students in the control group indicated decreased, i.e. 1.2 milliliter per kilogram body weight per minute or 3.9 percent (see Table 2).

Table 2. Fitness Score of Respondents in the Experiment and Control Groups

Fitness (VO ₂ max) ml/ kg body weight per minute	Experiment group (n=39) Mean \pm SD	Control group (n=40) Mean \pm SD
Mean \pm SD at Pretest	23.10 \pm 1.81	31.7 \pm 6.75
Mean \pm SD at Posttest	26.51 \pm 2.39	30.45 \pm 6.51
Mean different of Pre-Posttest \pm SD	3.41 \pm 2.33	-1.25 \pm 9.80
% Mean different of Pre-Posttest	14.76	-3.94

C. Academic Grade

Academic grades of the students in both groups were improved at the post test. Nonetheless, the improvement in the experiment group was higher than the control group. The academic grade of the students in the experiment group increased 26.6 or 60.3 percent meanwhile in the control group increased 3.3 or 9.6 percent (see Table 3).

Table 3. Academic Grade of Respondents in The Experiment and Control groups

Academic grade	Experiment group (n=39) Mean \pm SD	Control group (n=40) Mean \pm SD
Mean \pm SD at Pretest	44.15 \pm 11.71	34.12 \pm 6.21
Mean \pm SD at Posttest	70.76 \pm 4.72	37.39 \pm 9.30
Mean different of Pre-Posttest \pm SD	26.61 \pm 9.68	3.26 \pm 11.99
% Mean different of Pre-Posttest	60.27	9.57

D. Improved on Physical Fitness and Academic Grade in Each Group

Paired T-Test was used to know the different score of pretest and posttest on physical fitness and academic grade of the students in each group. At the experiment group, there were significantly improved on physical fitness and academic grade (*p value* 0.000). At the control group, indicated there were no significantly different score were found between pretest and posttest on physical fitness and academic grade (*p value* 0.425 and *p value* 0.093). See Table 4.

Table 4. Pretest-Posttest Dependent Variabels at The Experiment and Control Groups

Independen Variabel		Experiment group (n=39)		Control group (n=40)	
		Pretest	Posttest	Pretest	Posttest
Fitness	Mean±SD	23.11±1.89	26.52±2.39	31.7±6.75	30.45±6.51
	p value ^a	0,000*		0,425	
Academic grade	Mean±SD	44.15±11.2	70.47±4.73	34.12±6.21	37.39±9.30
	p value ^a	0.000		0.093	

^a Paired t-test*E. Affect Aerobic Gymnastic on Fitness and Academic Grades*

The data was analyzed by using Multivariate Analysis of Covariance after knowing that the data was normally distributed (p value more than or same with 0.05), with controlling age, gender, and parental education. Normality test showed that all data was normal distribution not only the physical fitness and academic grade of students at the experiment group but also at the control group. The results of normality test revealed p value 0.056 and 0.699 for the physical fitness and p value 0.499 and 0.958 for the academic grades.

The results of Multivariate Analysis of Covariance test showed that there were significant different independent variables (physical fitness and academic grade) between experiment and control group (p value 0.000). The significant different in detail on physical fitness and academic grade were p value 0.013 and p value 0.000, respectively after controlling age, gender and parental education. Test between subject effects indicated there was one co-variance affected the fitness, i.e. gender (p value 0.001) and there were two co-variances affected the academic grade i.e. gender and mother's education (p value 0.047 and p value 0.038). However, the effects of co-variances had been controlled. (see Table 5).

Table 5. Test of Between Subject Effect Co-Variances

Co-variables	Dependent Variabels-subject	Mean	F	p-value
Gender	Fitness	562.43	12.62	0.001*
	Academic grade	423.31	4,07	0.047*
Age	Fitness	31.57	0.70	0.403
	Academic grade	19.84	0.19	0.663
Father's education	Fitness	1.86	0.04	0.838
	Academic grade	336.75	3.24	0.076
Mother's education	Fitness	17.04	0.38	0.538
	Academic grade	463.72	4.46	0.038*
Experiment-control	Fitness	287.21	6.44	0.013*
	Academic grade	12152.06	117.11	0.000*

*significant (<0,05)

IV. DISCUSSION

A. Aerobic Gymnastic and Physical Fitness

Respondent's participation in aerobic gymnastic was influencing significantly their fitness as measured by maximum oxygen volume. Enhancement of their fitness was caused by its effect to lung capacity⁽¹⁰⁾, to heart-lung endurance⁽¹¹⁾, strength and endurance of body muscle including muscle of stomach⁽¹²⁾.

Enhancement of physical fitness of the respondents, whose age is around 20 was caused by participating in aerobic gymnastic for 4 weeks. This participation time was less than the result of Buchan DS research of the young adults i.e. 7 weeks⁽⁴⁾. Enhancement of physical fitness among 57-85 years old people, needed longer time than that of teenager's i.e. 12 weeks if it was expressed by lung capacity⁽⁸⁾ and needed 16 weeks if it was expressed by lung function⁽¹³⁾.

There was increasing score in academic grade of subjects in experiment group after they followed the aerobic session at 40 minutes duration, 5 days in a week in one month as much as 60.29 percent. This result was still in the percentage range of the previous result i.e. 55-68 percent (researcher from Dartmouth published by Huffington Post in Martin, 2010)⁽¹⁴⁾.

The increasing of the academic grade among experiment subjects was caused by increasing blood flow to dorsolateral prefrontal cortex. This condition indeed make physical activities as

aerobic gymnastic can show specific alteration at certain section of brain^(15,10). Afterward the nor-epinephrine and endorphins hormones will increase as well. This increases levels of nor-epinephrine and endorphins to decrease stress and improve mood⁽¹⁶⁾, or stress brain booster⁽¹⁰⁾. They will affect the brain growth i.e. growth of new nerve cell and support the connections between brain cell synapses, that are the basis of learning⁽¹⁶⁾.

Furthermore, with physical fitness expressed by healthy of heart and lungs (maximum oxygen volume) related to the efficiency of cognitive function⁽⁵⁾. In other words, physical fitness is needed to improve brain function. Ploughman says that exercise is brain food intake⁽²⁾.

The different in gender and mother education contributes to the escalating of academic grade. This finding parallels with result of the study of Pakistani medical students. This study showed there was correlated between gender and academic grade that measured by grade point average score. The score of male students were significantly higher than female ones⁽¹⁷⁾. However, the study of India's students in Uttar Pradesh showed the opposite results i.e. academic performance of male students wasn't significantly different with that of the female⁽¹⁸⁾.

Mother's education is one of the co-variances in this research. It has contribution on elevating the academic grade. This finding parallels with the research decade statement of Michigan Department of Education (2001). This statement was the earlier the mother involvement in their children education process, the stronger the effects on their children achievement⁽¹⁹⁾. Beside this, young mothers with low level of education will have children with high achievement if they provide supplementary education. However, mother education cannot predict their children achievement⁽²⁰⁾.

V. STUDY LIMITATION

Specific instrument for measuring academic grade was not developed and the time needed for intervention was restricted because the respondents ought to follow the final exam semester II as arranged by the academic calendar.

VI. CONCLUSION AND SUGGESTION

The partaking students of diploma 3 study program from the remote, border, and archipelago areas in North Kalimantan Province in Indonesia in aerobic gymnastic can significantly affected the physical fitness and academic grade. The aerobics gymnastics as sport can be used as an alternative to improve academic grade of students by maintaining their body fitness.

ACKNOWLEDGEMENT

Thank to Director of The Polytechnic of Health of Yogyakarta for assistance in funding the study, to lecturers of Human Basis Needs Subject in Diploma 3 Study Program both in Nursing Department and Dental Nursing Department for providing data of mid-term test and final-term test of semester II of the subjects; and to Ms Keni who helped this study as the aerobics instructor.

REFERENCES

- [1] Sibley, B.A; Etnier J. The relationship between physical activity and cognition in children: a meta-analysis. *Pediatr Exerc Sci.* 2003;15:243–56.
- [2] Plougman M. Exercise is brain food: the effects of physical activity on cognition function. *Dev Neurorehabil.* 2008;11:236–40.
- [3] Hartmann, T; Zahner,L; Puhse,U; Puder,J.J; Kriemler S. Effects of a school-based physical activity program on physical and psychosocial quality of life in elementary school children. *Pediatr Exerc Sci.* 2010;22:511–22.
- [4] Buchan, D.S; Ollis, S; Thomas, N.E; Buchanan, N; Cooper, S.M; Malina, R.M; Baker J. Physical fitness in childhood and adolescence: a powerful maker on health. *Int J Obes.* 2011;32:1–11.
- [5] 5Bherer, Louis; Erickson; Kirk I and L-AT. A review of the effects of physical activity and exercise on cognitive and brain function. *J Aging Res.* 2013;
- [6] Spirduso,W; Poon, L; and Chodzo-Zajko W. Using resources and reverse in an exercise-cognition model. *Hum Kinet Champaign Ill, USA.* 2008;3-11.
- [7] Direktorat Jendral Bina Gizi dan Kesehatan Ibu dan Anak KR. Kartu menuju sehat. In: Leaflet. 2011.
- [8] Renaud M. The effect of three month of aerobic training on response preparation in older adults. *Front Aging Neurosciences.* 2010;2:148.
- [9] Martin K. Brain boost : Sport and physical activity enhance children ' s learning The link between physical activity and academic interest in the field of education and. *Aust Gov West Aust.* 2010;(May).

-
- [10] Telles, S; Singh, N; Kumar Bardawa, A; Kumar, A; Balkrishna A. Effect of yoga on physical, cognitive and emotional measure in children: a randomized controlled trial. *Child Adolesc Psychiatry Ment Heal* (open acces Artic [Internet]. 2013;7:37. Available from: <http://creativecommons.org/licenses/by/2.0>
 - [11] Pennington, R; and Hanna S. The accute effects of exercise on cognitive performance of older adults. *J Ark Acad Sci.* 2013;67:109–14.
 - [12] Dwyer, T; Sallis, J.F; Blizzard, L; Lazarus, R; Dean K. Relation of academic performance to physical activity and fitness in children. *Pediatr Exerc Sci.* 2001;13:225–37.
 - [13] Dustman, R.E; Ruhlning, R.O; Russel E. Aerobic exercise training and improved neuropsychological function of older individuals. *Neurobiol Aging.* 1984;15(1):35–42.
 - [14] Karen M. Braint boost: sport and physical activity enhance children's learning. The University of Westter Australia; 2010.
 - [15] Cohen,D.L; Wintering,N; Tolles, V; Townsend, R.E; Farrar, J.T; Galantino, M.L; Newberg A. Cerebral blood flow effects of yoga training: preliminary evaluation of 4 cases. *J Altern Complement Med.* 2009;15:9–14.
 - [16] Paton G. Daily exercise signifcantly improve pupils test score. 2012; Available from: <http://www.telegraph.co.uk/education/educationnews/9133534/Daily-exercise-significantly-improves-pupils-tes>
 - [17] Farogi,Y.R; Ghani,R; Spielberger C. Gender differences in test anxiety and academic performance of medical students. *Int J Psychol Behav Sci.* 2012;2(2):38–43.
 - [18] Gupta,R; Sharma, S; Gupta M. A study of gender difference on the measure of academic achievement in adolescent students. *VSRD Tech &Non-technical J.* 2012;3(1):23–7.
 - [19] Michigan Department of Education Decision Making Yardistick. What research says about parent involvement in children's education. 2001.
 - [20] Magnuson K. Maternal education and child's academic achievement during middle childhood. *Dev Psychol.* 2007;43(1):1497–512.

Analyzing the authenticity of authentic assessment

Luki Yunita, Salamah Agung, Eka Novi

Program Studi Pendidikan Kimia
Fakultas Ilmu Tarbiyah dan Keguruan
UIN Syarif Hidayatullah Jakarta

Email: luki.yunita@uinjkt.ac.id; salamah.agung@uinjkt.ac.id; beti.chemistry10@gmail.com

Abstract- The 2013 Curriculum of Indonesia asks teachers to use an authentic assessment to evaluate their students. While such assessment has been practiced by teachers, study about students' perception on such assessment is really conducted. This study tries to examine high school students' perception on the authentic assessment according to five-dimensional framework of authentic assessment, namely, task, physical context, social context, form, and criteria. The questionnaire contained 40 items using Likert Scale that range from Always (4) to Never (0). Six public senior high schools in East Jakarta that were assigned to be pilot project for the implementation of the 2013 Curriculum participated in this study. One hundred ninety one high school students from the six schools completed the questionnaire. Results revealed that the authenticity of three dimensions of the assessment, i.e. the physical context, the social context, and criteria of assessment were considered to be adequate. The other two dimension, namely, task and form, on the other hand were found to be inadequate. This result serve as a stepping stone for the improvement of the practice of authentic assessment starting from its development and usage by the teachers.

Key Words: *Authentic Assessment, Perception, High School, Student*

I. INTRODUCTION

The curriculum 2013 has shifted the role of assessment, from a test assessment (measuring the competence of knowledge on by cognitive test) to an authentic assessment (measuring the students' attitudes, skills and knowledge based on the process and results). Authentic assessment is expected to facilitate students to use their competencies to solve daily problem. By using an authentic assessment, the objective of the 2013 curriculum, that is to create a person who has the ability to live as individuals and citizens who are productive, creative, innovative, and affective and able to contribute to the society, nation, state, and the civilization of the world.

Assessment in the curriculum in 2013 refers to the Minister of Education and Culture of the Republic of Indonesia Number 66 Year 2013 about the Standard Assessment of Education which aims to ensure: Planning students' assessment in accordance with the competencies to be achieved and based on the principles of valuation; implementation of students' assessment in a professional, open, educational, effective, efficient, and in accordance with the socio-cultural context; and reporting students' assessment result in an objective, accountable and informative. Educational assessment as a process of collecting and processing information to measure the achievement of learning outcomes of students is including authentic assessment, self-assessment, portfolio-based assessment, quizzes, daily tests, midterm test, final test, the national exam, and school/madrasah exam.

The shift in the assessment has caused some constraints in practical level. Based on preliminary research through interviews with teachers in schools in which the 2013 curriculum was piloted, it was found that teachers faced difficulties in implementing the assessment plan and process. In the assessment process, for example, teachers were required to give a complete picture of the assessed attitudes, skills and knowledge of students associated with their real life outside the school or community. Furthermore, they were also asked to create appropriate assessment format. However, the interview results revealed that teachers found it difficult to do such assessment. This is due to the lack of professional training. If such assessment is continued to be used and implemented by teachers whose knowledge are lacking, the goal of authentic assessment might not be well achieved and thus the educational objectives of 2013 curriculum might not be met.

A good assessment is expected to have a positive influence, particularly, on the way students learn. The positive influence consists of two things, namely, stimulating the development of professional competence of students and increasing students' motivation to learn. The later required assessment that is more interesting, meaningful, and is linked with their daily experiences (Judith M. Gulikers, 2008: 74). Students as learners have an important position in the assessment process. When learners are involved in the assessment, they can learn more effectively. , and there is a possibility that they can adjust to the standards that teachers use assessment. If the students were asked to reflect on and evaluate their own performance, then most likely they are more motivated to learn. Thus, learning outcome might be easily achieved.

An assessment that will provide a good result should have been understood by students. Yet, most of the time, students do not know and understand what being assessed and how their grade is finalized. In terms of authentic assessment, students' dan teachers' understand on such assessment might be different. Teachers may think that they have carried out an authentic assessment. Students, on the other hand, might not realize that they have been assessed using an authentic assessment. Such difference in perception, according to Entwistle (2012), was caused by different experiences, motivations, and desires of students and teachers. Furthermore, Honebein , Duffy , and Fishman as cited in Gulikers (2008) mention that different perception on the assessment may be caused by different of age, level of education, and experience. However, since the main target of authentic assessment is student, the need of students to understand and to have the same perception as teachers about the authentic assessment is of very important. Once students and teachers are in agreement with the assessment, the objective of learning will be well achieved.

Referring to the urgency of implementing authentic assessment for education in Indonesia as well as the difficulties and constraints in implementing it, it is necessary to investigate whether the application of authentic assessment done by teachers has been going well in accordance with the principles of authentic assessment. Futhermore, it is also necessary to know whether there is a discrepancy of perception of authentic assessment hold by students and teachers. To that end, the authors conducted a study to analyze the authenticity of authentic assessment conducted by teachers in schools based on student perceptions. A questionnaire to measure students' perceptions is developed based on five dimensions of authenticity introduced by Gulikers , Bastiaens, and Kirschner. The five dimensions include tasks, physical context, social context, forms, and criteria.

II. METHOD

This research was conducted in SMA Negeri 48, SMA Negeri 54, SMA Negeri 58, SMA Negeri 81, SMA Negeri 100, and SMA Negeri 113 of East Jakarta. This research was a quantitative descriptive research that tried to describe the result of a survey (Nana, 2011:72). The survey questionnaire included five dimension of authenticity, namely, tasks, physical context, social context, forms, and criteria. The survey was considered to be a perception survey in which students were asked their perception of the five dimensions. The questions were constructed with a Likert scale from 0 to 4 (never, almost never, rarely, sometime, always).

III. RESULT AND DISCUSSION

Graphic 1 shows the percentage of students' perception on the five dimensions of authenticity of authentic assessment. According to the graphic, in general, students think that they have been assessed using authentic assessment (above 70%). Among the dimensions, the task dimensions and form dimensions are indicated to have low percentage. The former defines the task itself while the later defines the form of task students are assessed. Physical context is considered to be the highest percentage. This dimension indicates the accuracy/timeliness/tools/materials and professional resources that are used in the assessment process is already authentic. Social context dimension is also indicated to be well perceived by the students. This dimension determines interaction of students during the assessment. While indicating positive perceptions of students, these results are not in accordance with the result of Gao (). Gao finds that 78% of respondent do not feel that they are assessed with authentic assessment. They feel almost never have questions on the relation of theory with practice (2012). This indicates that the form of assessment is not linear with authentic assessment dimensions.

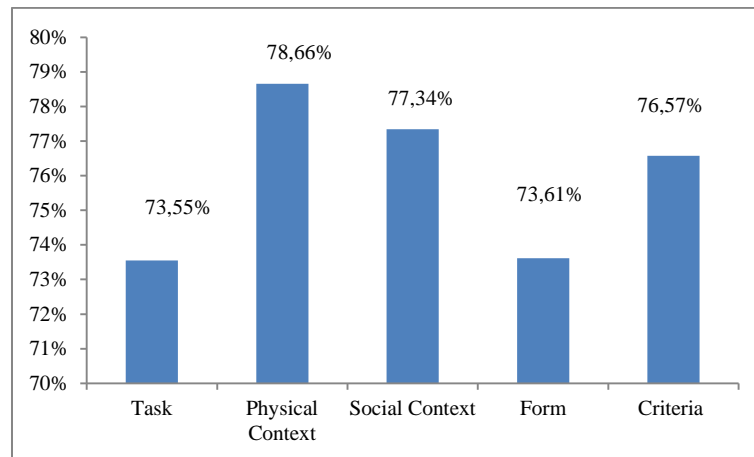


Figure 1. Grafik Persentase Persepsi Siswa Tentang Autentitas Penilaian Penilaian Autentik Berdasarkan Lima

IV. SUMMARY

According to the results, it can be described that students perceived their assessment as authentic assessment. In other words, the authenticity of authentic assessment has been well applied in the assessment process.

REFERENCES

- [1] [1] Anisa, Alita Arifiana. "Evaluasi Penerapan Penilaian Otentik dalam Kaitannya dengan Kesiapan SDM Menghadapi MEA". (*Evaluation on the implementation of authentic assessment in relation to the readiness of human resources facing MEA*) Prosiding Seminar Nasional Penelitian dan Evaluasi Pendidikan. 9 Mei. Yogyakarta: Universitas Negeri Yogyakarta, 2015.
- [2] Entwistle, Noel. *Styles of Learning and Teaching, An Integrated Outline of Educational Psychology for Students, Teacher, and Lecturers*. New York: Routledge, 2012.
- [3] Gao, Minghui. *Classroom Assessments in Mathematics: High School Students' Perceptions*. International Journal of Business and Social Science. 3, 2012.
- [4] Gulikers, J. T. M., et al. *A Five-Dimensional Framework for Authentic Assessment*. Educational Technology Research and Development, 52, 2004.
- [5] Kementerian Pendidikan dan Kebudayaan Republik Indonesia. "Salinan Lampiran Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 66 Tahun 2013 tentang Standar Penilaian Pendidikan", <https://www.kemdikbud.go.id>, 7 Desember 2014.
- [6] Kunandar. *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013), Suatu Pendekatan Praktis Disertai dengan Contoh. (Assessing students performance based on 2013 curriculum – a practical approach with examples)* Jakarta: PT RajaGrafindo Persada, 2013.
- [7] Miller, M. David, et al. *Measurement and Assessment in Teaching*. New Jersey: Pearson Education, 2009.
- [8] Muslich, Masnur. *Authentic Assessment: Penilaian Berbasis Kelas dan Kompetensi*. Bandung: PT Refika Aditama, 2011.
- [9] Mussawy, Sayed Ahmad Javid. "Assessment Practices: Students' and Teachers' Perceptions of Classroom Assessment", Tesis pada University of Massachusetts Amherst: 2009. tidak dipublikasikan.
- [10] Riduwan. *Belajar Mudah Penelitian Untuk Guru-Karyawan dan Peneliti Pemula (Easy learning research methodology for teachers, professionals, and researchers)*. Bandung: Alfabeta, 2013.

DESIGN OF PERFORMANCE ASSESSMENT BASED ON PROBLEM BASED LEARNING IN IMPROVING STUDENTS' SELF REGULATION

Luthfi Riadina¹, Agus Suyatna², Undang Rosidin³

¹Teacher in Junior High School Satu Atap 9 Pesawaran,

²Master of Physical Education Guidance and Counseling University of Lampung

³Master of Physical Education Guidance and Counseling University of Lampung

¹luthfiriadina@rocketmail.com

Abstract - The purposes of this research are: 1) understanding the needs of teachers and students will be performance assessment that can help in improving the skills of self-regulation of students, 2) analyze the learning process of physics in several high schools in the District Pesawaran, 3) designing the instrument performance assessment-based problem-based learning (PBL). Preliminary studies in this study consisted of library research and field study. Studies conducted in the literature study are to learn the concepts related to the study. While a field study conducted by distributing questionnaires to several teachers and senior high school students in the District Pesawaran. The subjects were a few teachers and students of senior high school in the county Pesawaran. Data were analyzed qualitatively through four stages: encoding the results of questionnaires, data tabulation, analysis of qualitative data, and make interpretation of results of analysis in accordance with the research questions and make their conclusions. The findings of this preliminary study are: 1) Teachers and students in several senior high school in the county Pesawaran need an instrument performance assessment to help students develop skills of self-regulation, 2) Some physics teacher senior high school in the District Pesawaran already know and never apply Problem Based Learning in learning but still feel difficulty in its use; 3) Design a performance assessment instrument developed tailored to the stage of Problem Based Learning in order to improve students' Self Regulation.

Keywords: *performance assessment, PBL, self-regulation*

I. INTRODUCTION

Learning is the most important activity in education and a systematic program with a specific sequence of steps starting from planning, implementation to evaluation/ assessment. The quality of learning is determined by the level of mastery of concepts students to the material being studied. Evaluation/ assessment is one of the lessons that cannot be separated from the learning itself. Assessment is particularly important as the assessment can provide information to teachers about student progress in the learning process. [1] assessment provides information that can help improve students' learning and help teachers in teaching. The performance assessment is one form of assessment is considered accurate enough to determine student learning outcomes. Performance assessment involves students in activities that require demonstration of certain skills and/ or in terms of creating a product specification. Assessment of performance is very close relation to the ongoing process of learning that results in students' understanding of the material presented.

The performance assessment will provide an opportunity for students to demonstrate their learning ability in its totality from start to finish that can be evaluated by others in this case the teacher or lecturer [2]. Performance assessment involves students to demonstrate their ability to think, to do a specific skill, and / or to create a particular product [3]. [4] Assessment is a component of the learning process traditionally has been accomplished through pencil and paper. The performance assessment is an approach that is needed to assess the learning outcomes of gifted students [5].

To improve thinking skills in mastering a concept needs to use proper methods and techniques in the delivery of the material. The relationship between teaching and learning models are closely related, because the use of appropriate learning models will produce a good learning process as well.

To provide feedback so that students are able to develop the capacity to think, need to use appropriate learning techniques. Problem based learning (PBL) can help students to develop thinking skills and improve student learning independence. [6] States Problem Based Learning, will help students to develop their cognitive, instrumental and transversal competences, which will allow them to expand their learning opportunities. Not only that, PBL will also make students more active in seeking a solution, solving a problem, and will be more active in track down something related to his problem. [7] States that PBL increases students' thinking skills, and will also provide a permanent learning, as well as active participation in learning, attracts the attention and interest of the students to learn. Problem Based Learning leads to better test performance with minimal mental effort invested in studying and learning more efficient[8]. The resulting performance on the learning problem based learning will foster liveliness, interest and awareness of students in learning physics.

Self-regulation (SR) is a process of self-assessment in which students represent tasks, plan how to implement, monitor and assess whether the implementation is sufficient, able to overcome difficulties and emotional that normally arise, assess performance and make the relationship conclusions about the results [9]. Self-regulation is also interpreted as a self-conscious process on psychic activity that ensures the setting as well as progress in meeting the specific purpose [10]. If the process is self-aware students will learn to high, then the students can learn the maximum. Therefore, self-regulation is considered important and is needed by every student.

II. METHODS

This research is preliminary. In this study generated assessment performance-based design problem based learning (PBL) in improving students' self regulation. The sampling technique is cluster random sampling means that sample by selecting one or several groups by simple random sampling as a sample. The subjects were students of class X Senior High School totaling 20 students and 4 teachers physics 2015/2016 school year in some schools in the District Pesawaran.

The data collection is done 2 steps, the first step in analyzing the needs that are literature study and field study. Studies conducted in the literature study is to learn concepts related to the study. While a field study conducted by distributing questionnaires to several teachers and high school students in the District Pesawaran.

Analysis of the results of questionnaires carried out by to use several steps: encoding the results of questionnaires that have been distributed, the tabulation of the data makes it easier to read, analyze qualitative data, and makes interpretation in accordance with the results and make conclusions.

III. RESULTS AND DISCUSSION

Based on analysis of the requirements found in the field, it is known that teachers need effective instruments of performance assessment that can be used to enhance the students' self regulation.

Table 1. Analysis of the implementation of the Performance assessment by teacher

No	Analysis of Answers
1.	As many as 75% of high school physics teachers have been to use performance assessment of learning
2.	A total of 100% high school teacher considers it important performance assessment
3.	A total of 100% high school teachers feel the need for performance assessment instruments
4.	As many as 75% of high school teachers find it difficult to use the instrument of performance appraisal

Based on the results of the analysis showed that the majority of high school teachers in the district of Pesawaran aware of the importance of performance assessment and they have been to use the instrument performance assessment for learning. However, they find it difficult to use

the instrument. So most teachers feel the need for a performance assessment instrument that is easily understood and used by teachers. The analysis in the field also showed 75% of teachers agree if developed performance assessment instrument to facilitate teachers in making judgments.

Table 2. Analysis of the implementation of the Performance assessment by students

No.	Analysis of Answers
1.	45% of students said the teacher of physics tell about the aspects to be assessed and how the assessment to all students at the beginning of learning
2.	60% of students said the teacher of physics are often assessed when the learning process
3.	60% of students said teachers often observe physics student activities during the learning process
4.	30% of students say teachers of physics ever conduct performance appraisals that were previously communicated to students
5.	70% of students to be more active when assessing the performance of the previously communicated to students
6.	50% of students have been asked by his physics teacher to demonstrate the performance and skill in learning and then assessed by teachers
7.	45% of students feel less fair with the assessment given by his physics teacher
8.	60% of students had shown good performance during the learning process to get good grades in physics

Based on the analysis of the answers obtained from the questionnaires given to students can be found in accordance with the statement of the number 4 only 30% of students say that teachers of physics ever performance appraisal. These statements contradict the statements of student numbers 2,3, and 6 which states that the majority of students said physics teacher they ever ask students to demonstrate its performance during the learning process and their physics teacher also often assessed by observing the activities and performance of students during the learning process ongoing. This was probably due to lack of understanding of students will be the performance appraisal itself. Meanwhile, the students feel would be more active if the assessment of performance previously communicated to students in advance. Hence, the explanation needs to be given to students on performance assessment.

Table 3. Analysis of Teacher Knowledge to be Self-Regulation

No	Analysis of Answers
1.	As many as 50% of teachers SMAN know about self-regulation
2.	A total of 100% high school teachers to motivate students to study harder
3.	As many as 75% of teachers consider the level of awareness of students in learning and learning outcomes

Based on this analysis well known that most teachers understand the self-regulation or the level of understanding and awareness of the importance of students' learning. Their self-regulation in students allows students independent learning. Students involved a learning environment that supports independent learning, have to manage their own learning and take responsibility for learning in other words, they must become independent learners [11]. Independent learning in education is based on the premise that students use metacognitive, motivation, and behavior in their learning process [12]. In this case, it takes a learning model that can enhance students' self regulation. It is based on the results of the analysis of the needs of students in the field.

Table 4. Analysis of the Needs of the Students will be Learning PBL

No	Analysis of Answers
1.	In the study of physics, 100% of students say that they have had in learning by groups
2.	In the study of physics, 70% of students say that the physics teacher never given a problem to solve in groups
3.	In the study of physics, 90% of students say that they are happy if learning is done in groups
4.	In the study of physics, 80% of students agreed when learning physics conducted in groups and presentations.

Based on the results of this analysis can be seen that most of the students prefer it if the learning is done in groups. Therefore, the model of Problem Based Learning (PBL) be the perfect

solution in improving students' self regulation. Problem Based Learning aims to develop student-centered environment [13].

Model PBL used by researchers to use the stages Furthermore, of the eight principles of PBL, Miller in [14] noted several strategies in PBL include: (1) learning begins with activities and simulations, (2) a demonstration based on the data, (3) discussion group, (3) individual and group projects, (4) The written and oral presentations, (5) the presentation of the material-based activities, (6) make predictions before carrying out the activity, (7) pushed to do a lot of representations, think, and reflect , and interact with peers.

It is known that most of the students prefer it if the learning is done in groups. Therefore, the model of Problem Based Learning (PBL) be the perfect solution in improving students' self regulation. Problem Based Learning aims to develop student-centered environment [15].

Model PBL used by researchers to use the stages Furthermore, of the eight principles of PBL, [16] noted several strategies in PBL include: (1) learning begins with activities and simulations, (2) a demonstration based on the data, (3) discussion group, (3) individual and group projects, (4) The written and oral presentations, (5) the presentation of the material-based activities, (6) make predictions before carrying out the activity, (7) pushed to do a lot of representations, think, and reflect , and interact with peers.

Design performance assessment wants researchers to develop other than by to use problem-based learning, the researchers also wanted to improve the existing self-regulation in students. Through Problem Based Learning students will become more active in its performance. Thus the awareness of students to study physics will grow not only improve on the affective aspect but also the cognitive and psikomotornya. The stages of problem based learning that researchers developed in this study were: 1) the orientation of students on the issue; 2) Discussion Group; 3) The investigation / search for solutions; 4) present the results of the discussion; and 5) the analysis and evaluation of problem-solving process.

Table 5. Performance Assessment Instrument Design

Stages	Implementation Techniques	Instrument
Orientation	Observance	Observation sheet
Group discussion	Observance	Observation sheet
Search Solutions	Written test	Problem Description
Presentation	Observance	Observation sheet
Analysis	Written test	Problem Description

IV. CONCLUSION

Based on the above research purposes, it can be concluded: 1) Teachers and students in need of a performance assessment instrument that can assist students in developing self-regulation; 2) Some high school physics teacher in the District Pesawaran already know and never apply Problem Based Learning in learning, but still had difficulty in its use; 3) Design a performance assessment instrument developed tailored to the stage of Problem Based Learning to help increase students' Self Regulation.

ACKNOWLEDGMENT

This project would not be possible without the help from members of our team who are not author on this paper Erlida Amnie, Asih Sulistia Ningrum, Feryco Chandra, Saiful Ali Nurdin, Imas Setana Esti Galih, Nurul Ulil Amri, Astri Mela Agustin, Tuti Widyawati, Novinta Nurulsari, F. Bayu Nirwana, Yuda Seta Mahendra And Muhammad Iwan.

REFERENCES

- [1] Kerry Earl, and David Giles, "An-Other Look at Assessment: Assessment in Learning," New Zealand Journal of Teachers' Work. New Zealand, Vol. 8, pp. 11-20, 2011.
- [2] Ferdy Dungus, "Thde Effect of implementation of performance assessment, portofolio assessment and written assment toward the improving of basic physics II learning achievement", Journal of education and practice.Indonesia, Vol.4, pp 2222-288X No.12013.
- [3] M. Metin, and H. Ozmen, "Investigation of Teacher Opinions about Performance Assessment with Respect to the Gender and Branch Variables," Journal of turkish science education.Turki, Vol.8 pp 3-17, 2011.
- [4] Melissa H Dancy, "Impact of animation on assessment of conceptual understanding inphysics", Physics education research 2. ISSN: 010104,2006.
- [5] J.Baska, and Van Tassel, "Performance Based Assessment The Road to Authentic Learning for the Gifted", Journals permissions. Vol 37, 2013.

- [6] L. M. Draghicescu, Ana-Maia Petrescu, G.C. Cristea, L.M. Gohiu, and Gabriel Groghiu, "Application of Problem Based Learning Strategy in Sains Lessson Examples of Good Prractice", *Procedia. Romania*, pp 1877-0428.2014.
- [7] A.rzu Ari, and Yasemin Katranci, "The opinions of primary mathematics student teachers on problembased learning method," *Procedia*.ISSN:1826-1831. 2013.
- [8] Noor Hisham Jalani and Lai Cheei Sern, "Efficiency Comparisons between Example-Problem-Based Learning and Teacher-Centered Learning in the Teaching of Circuit Theory," *Procedia. Malaysia*, Vol.204, PP 153-163, 2015.
- [9] E. Panadro, J.A. Tapia, and J.A. Huertas, "Rubrics and Self-Assessment Scripts Effects on Self-Regulation, Learning and Self-Efficacy in Secondary Education," *Learning and Individual Differences*, Vol. 22, pp 806-813, 2012.
- [10] Morosanova. "Self regulation and Personality," *Procedia. Russia*, Vol. 86 pp 452-457, 2013.
- [11] M. Ekici, H.I. Coskun, and H. Yurdugul, "Investigation of the relationship between learning approaches and online self-regulation behaviour," *Procedia. Turkey*, Vol. 141 pp 285-289, 2014.
- [12] *De Fatima Maria Goulao, and Menedes Rebeca Cerezo*, "Learner autonomy and self-regulation in eLearning," *Procedia*.ISSN:1900-1907, 2015.
- [13] Kawita Panlumlars, and Panita Wannapiroon, "Design of cooperative problem based learning activities to enhance cooperation skill in online environment," *Procedia. ISSN:2184-219*, 2015.
- [14] Stanley P Dewanto, "Peranan Kemampuan Akademik Awal, Self-Efficacy, dan Variabel Nonkognitif Lain Terhadap Pencapaian Kemampuan Representasi Multipel Matematis Mahasiswa Melalui Pembelajaran Berbasis Masalah," *Educationist. Indonesia*, Vol. II ISSN: 1907 – 8838, 2008.
- [15] Kawita Panlumlars, and Panita Wannapiroon, "Design of cooperative problem based learning activities to enhance cooperation skill in online environment," *Procedia. ISSN:2184-219*, 2015.
- [16] Stanley P Dewanto, "Peranan Kemampuan Akademik Awal, Self-Efficacy, dan Variabel Nonkognitif Lain Terhadap Pencapaian Kemampuan Representasi Multipel Matematis Mahasiswa Melalui Pembelajaran Berbasis Masalah," *Educationist. Indonesia*, Vol. II ISSN: 1907 – 8838, 2008.

IMPLEMENTATION OF PERFORMANCE ASSESSMENT TO INCREASE BIOLOGY LEARNING ACHIEVEMENT BY USING INQUIRY MODEL

Murni Sapta Sari

Biology Department Faculty of Mathematics and Science Malang State University
murni.sapta.fmipa@um.ac.id

Abstract - This study is conducted in order to understand the performance assessment implementation to increase Biology learning achievement by using inquiry model. The Study used action research type- based lesson study. Lesson study stages included planning, doing, and observing. The action reseach was conducted in two cycles. Each cycle was performed in three meetings. Data collection process started in 16 May until 26 September 2015. The subjects of the study consisted of students XI MIA 5 of High School 1 Lawang, periode of 2015/2016 comprising 14 male students and 19 female students. The instruments of data collection were performance assessment instruments consisting of observation rubric, self assessment rubric, and learning achievement test. The result of the research shows that the implementation of performance assessment by using inquiry model can increase biology learning achievement of students XI MIA 5 of High School 1 Lawang. Beside, the result of the research can be used as a reference for teachers in conducting better learning process by lesson study.

Keywords: inquiry model, lesson study, performance assessment

I. INTRODUCTION

The curriculum of 2013 applies performance assessment which is expected to result in qualified Indonesian citizen who are productive, creative, inovative, and affective through integrated empowerment on attitude, knowledge, and skill. Student's learning achievement on the three aspects can be assessed by various methods; one of them is performance assessment, performance assessment can be used to measure knowledge, skills, and attitudes^[1]. Performance assessment has a number of strengths, namely, 1) assessment that can make the students apply answers or results through demonstrating or performing their knowledge, skills or performance, 2) motivating the students during learning in a more sufficient way, 3) Encouraging application of learning outcome into the real situation. This is because performance assessment highlights on what the students can do, but not on what is known^[2].

Assessment is often considered as one of three main principles determining instructional activities. The three principles are, planning, implementation, and assessment. If the three principles work synergetically and sustainably, this will heavily determine the instruction quality. Therefore, the assessment must be designed and implemented according to instructional planning and implementation. Teachers should have changed their way of assessing the students, as matter of fact, they also needed to change the way they teach and the way the student learn^[3]. Performance is authentic assessment that can serve as effective media to measure complex skills or sort of skills which cannot be measured by paper and pencil based test. Measurable skills like planning, investigating, communicating, problem solving, and reasoning skills are included in inquiry stage^[4]. On the other hand, inquiry learning is defined as an active scientific exploration process instructing students to use their reasoning ability, logic, and creativity. Guided inquiry model has stages as follow: exploring a phenomenon, focusing on question, planning investigation, conducting an experiment, analyzing data, synthesizing new knowledge, and disseminating the new knowledge. The students are encouraged by their

curiosity to investigate towards phenomena they observe during instruction^[5]. The students, then, are able to spur questions or problems, to choose actions and to carry out investigation procedures, to collect and analyze data through observation activities and to derive a conclusion from the instruction in the class.

II. RESEARCH METHOD

The research methodology used for this study was a qualitative method with classroom action research. Classroom action research model^[6] applied two cycles, namely a spiral form from one cycle to the next cycle. Each cycle included three steps namely planning, action and observation, and reflection. The next step in the cycle comprised the already revised plan, action, observation, and reflection. Before implementing the first cycle, a preliminary action was taken by identifying problems. The research was conducted in the grade XI MIA 5 SMAN 1 Lawang with a total of 33 students in May - September 2015. Instrument that is used in the research is (1) Cognitive test problem to know students cognitive achievement, 2) learning observation sheets to measure learning activities, 3) Observation sheets to assess students' attitude and skill, 4) field note form to record unrecording activities in observation sheets.

The research on the implementation of performance assessment to increase the students' learning achievement was carried out within two cycles. Every cycle consisted of planning, action, observation, and reflection. This research was conducted and based on lesson study, comprising 3 stages, namely, plan, do and see. The instructional model conducted was a guided enquiry model. This is because the students were not yet familiar to do complete enquiry model. Therefore, the planning stage in the classroom action research was carried out collaboratively within a biology MGMP group in Malang. Meanwhile, 'do' and 'see' stages were held on the classroom action research stages, included in the observation and reflection. According to the observation activities held on the action stage, the result was obtained. Data collection from learning result consisted of aspects, namely, attitudes, skill performance on microscopic observation and delivery skills. In terms of delivery skills, knowledge aspect was also found and these entailed clarity in delivering new knowledge, the depth of content, and argumentation skills.

III. RESULT AND DISCUSSION

The observation and interview results from March to April 2014 in some senior high schools in Malang district found that 1) Teachers had carried out the instructional activities very well. The instructional activities designed by the teachers had already focused on discussion and practice method, but the process had not reflected holistic inquiry. For example, the activities was not initiated by brainstorming about everyday life issues required to be solved scientifically. Additionally, during laboratory activities on plant-cellular structure and network the teachers had not benefitted from the surrounding plants for microscopic observation, and the students were only limited to observe a preserved preparation. As a consequence, the students had not been able to value the given content, 2) In general, the assessments given by the teachers were less varied in terms of methods and instruments. The assessment was chiefly devoted to mastery of content which was used for an objective test serving for low level cognition like memorizing, 3) the teachers had not fully understood and carried out performance assessment. Based on the forementioned problems, it is necessary to have revision on instructional activities by implementing performance assessment using lesson-study-based inquiry model.

The research result showed an increase in students' learning and social attitudes from cycle I and cycle II as much as 15.25%. During action and observation stages depicted that the teachers in Senior High School of Malang Region had not been familiar to measure attitude aspect. Assessment involving cooperation, careful, and responsibility skill. The comparison of learning result, attitude aspect from cycle one and cycle two are presented below on Figure 1.

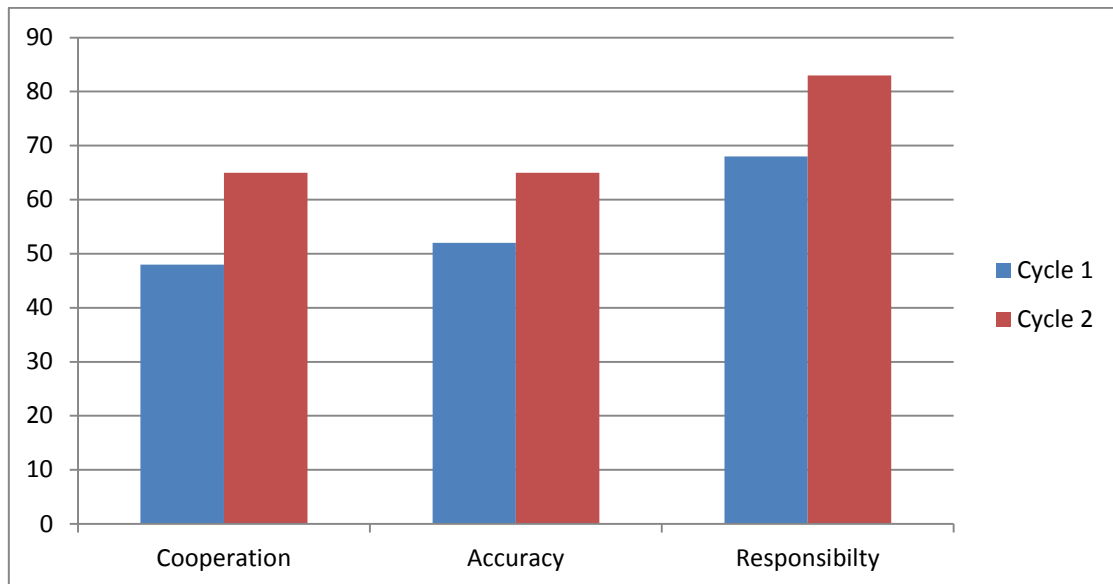


Figure 1. The comparison of social attitudes on cycle I and cycle II

Cognitive achievement assessment in cycle I determined based on test score. Based on assessment result known that 23 students success (69%), and 10 students (31%) failed. Success criteria for students is 75 grade, and success criteria for class is 85% students success in test. Assessment of skill that being measured is cells and tissues observation skill by using microscope and communicate the result. Microscopic observation skill involve: work safety, preparing tool and material, making sliced glass, operating microscope, drawing, accurate observation and keepcleanilness. Performance skills on microscopic observation was presented on Figure 2.

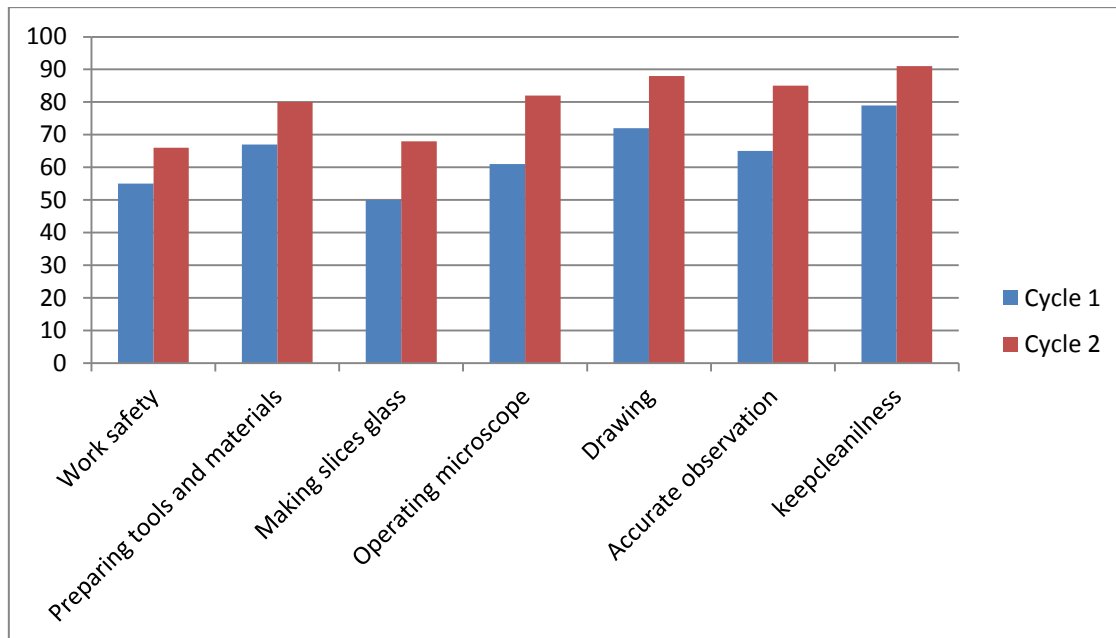


Figure 2. The Comparison of Performance Assessment on Microscopic Observation on cycle 1 and cycle 2

Communication skill involve clarity on new knowledge, content depth, argumentation skills respecting each other opinion, using Indonesian language.

The following performance assessment is commnucation skills on Figure 3.

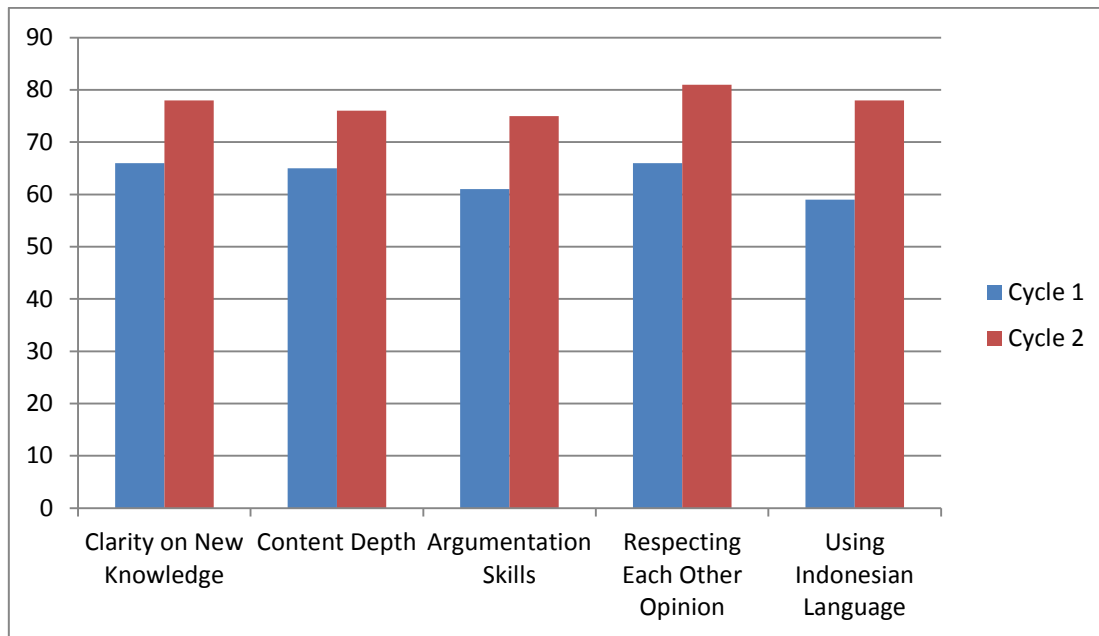


Figure 3. The comparison of communication skills

IV. DISCUSSION

The classroom action research conducted in XI MIA 5 SMAN Lawang based on lesson study was a program to develop teachers' professionalism. Besides, the implementation of 2013 curriculum expected teachers to be able to carry out duties professionally. However, the result indicated that the unfulfilled professional demand can be a major problem for the implementation of 2013 curriculum. In the planning activities there was a discussion on collaborative instruction by the members of MGMP group and the discussion results were met in which the chosen basic competency included Basic competency 3.1; 3.2; 3.3 and 3.4. The reason behind the chosen basic competency was that the complexity level it occupied was considered high to be implemented during teaching and learning activities. One of the reasons on why lesson study was essentially to be carried on was that the teacher would be able to translate basic competency collaboratively, especially the purpose and the standard of education which was turned into real class activities, initiated by planning activities, instructional instruments, Do (action), and see (observation, reflection toward observation results)^[7].

Inquiry model based on teacher-student dominance level. There are 4 types of demonstrated enquiry or discrepant events, structured inquiry, guided inquiry and full inquiry^[8]. The stages of guided inquiry are relevant to scientific approach in 2013 curriculum. During planning activities, apart from determining the instructional model, assessment method must be following. Teachers should change the instructional method by changing the assessment method because if the teachers change the way they assess students' learning, they also have to change the way they teach and the way the students learn^[3]. Based on that reason, therefore, performance assessment proposed by ^[1] was applicable to be implemented to measure knowledge aspect, skills, and attitudes.

In general, the problem relied on limited time to assess through observation sheet. As a consequence, it was quite often for the teachers to use questionnaires filled by the students. This would decrease the value of objectivity. Thus, observation sheet was needed. To add, attitude aspects measured also needed to be highlighted, in which they must be highly relevant to the content during instructional process. Moreover, the number of attitude aspects should not be too many. Competency assessment on attitudes is a series of activities designed to measure students' attitudes resulted from an instructional program. Likewise, attitude assessment is a kind of the implementation of an instructional standard or system of decision making towards attitudes. The main purpose of attitude assessment is a part of reflection and reasoning instruction, and student's individual progress toward attitudes.

Skill that being measured is microscopic observation skill and communication skill that is increased insignificantly that is only 15, 80%. Based on the research, known that learning

strategies with guided inquiry model can be used to measure and assess microscopic observation skill and communication skill in accordance with the opinion of ^[8]. Through exercise in inquiry learning that imitate the real scientific research in general and simple way, students will construct and develop their own understanding in the nature of science and scientific thinking skill as well as develop inquiry skill in the same time^[8].

In 2013 curriculum the students are advised not to be passively assisted during learning, but they are advised to do discovery, so that there was a need to facilitate classroom activities with guided inquiry model. This interesting finding found during lesson study activity especially on 'do' and 'see' stages was that the biology teachers in implementing inquiry model were not as complete as it was. So that the students could not develop their process of inquiry. The implementation of guided inquiry model of the first cycle according to teachers achieved 90%; it means that the quality of the instruction by the teachers was categorized good.

The result of the research show that guided inquiry model can increase students learning achievement. Learning based on inquiry is a learning approach that involve question and students investigation result of some problem as learning experience. Teacher must encourage students to state and investigate their ideas then communicate their investigating result. Inquiry learning model based on the ideas that is teacher and students have their own responsibility in learning activities^[9]. The advantages of guided inquiry model namely: (1) intellectual potentiation, students learn to develop their thinking if they use their thinking to think, (2) intrinsic motives, students accept intellectual vibration that satisfy an intrinsic vibration as consequence of guided inquiry, (3) students did guided inquiry if they have a chance to do it, and (4) memory saving, inquiry learning can help memory retention^[10].

The result of the study on cycle I and cycle II on delivery skill showed an increase at 14.48%. The aspects being measured were 1) clarity in delivery new knowledge, 2) the depth of the content, 3) the ability in argumentation, 4) respecting other people's opinion, 5) using Indonesian language the aforementioned five aspects being measured could not be valued using paper and pencil test so that performance assessment was needed. There are several reasons why performance assessment needs to be conducted in a class, those reasons are: 1) giving more opportunities to teachers to know the students even more holistically because; in fact, not all students who are less successful in an objective test or an open answer test are usually acknowledged as not skillful or not creative, 2) being able to witness the students' skills performed during the instruction process without waiting even until the lesson ends, 3) there are skills developed by the students which are not easily observable through solely referring to a written test or the final result of their assignment^[1].

V. CONCLUSION

The implementation of inquiry model of performance assessment could improve the learning result of students on grade XI MIA of SMAN 5 Lawang. The learning result on social attitudes increased from cycle I to cycle II. Moreover, the activity of lesson study was also found to improve the instruction quality through collaboratively and continuously discussing the instructional activities based on partnership principle and mutual learning, as well as establishing a learning community.

REFERENCE

- [1] Stiggins, R.J. 1994. *Student Centered Classroom Assessment*. New York: maxwell Macmillan International Simon & Schuster Company.
- [2] Airasian, P.W. 1994. *Classroom Assessment*. New York: McGraw- Hill, Inc.
- [3] Hart, D. 1994. *Authentic Assessment A Hand Book for Educators*. California, New York: Addison-Wesley Publishing Company
- [4] Marzano, R. J., 1993. *Designing a New Taxonomy of Educational Objectives*. Thousand Oaks, CA:Corwin Press
- [5] Llewellyn, D. 2013. *Teaching High School Science Through Inquiry and Argumentation*. USA: Crown.
- [6] Kemmis, S. and R McTaggart, 1988. *Action Research - some ideas from The Action Research Planner, Third edition*, ed. Deakin University
- [7] Susilo, H. 2009. *Lesson Study. Berbasis MGMP sebagai Sarana Pengembangan Keprofesionalan Guru*. Malang: Penerbit Surya Pena Gemilang
- [8] Ben-David, A., & Zohar, A., 2009. Contribution of meta-strategic knowledge to scientific inquiry learning. *International Journal of Science Education*, 31(12), 1657-1
- [9] Scardamalia, M., 2002. Collective cognitive responsibility for the advancement of knowledge. In B.Smith. *Liberal education in a knowledge society*(pp 67 – 98). Chicago, H. Open Court.

- [10] ¹Carin, A. A dan Sund, B.R. 1995. *Teaching Science Through Discovery*. Columbus: Charles E, Meril Publishing Co. A Bell & Howel Company.
- [11] Popham, W. James, 2011. *Classroom Assessment: What Teachers Need to Know*.
- [12] Needham Heights, MA; Allyn & Bacon, A. Simmon & Schuster Company.

TEACHERS' BELIEF IN IMPLEMENTING FEEDBACK FOR STUDENTS' WRITING IN ESP CLASSROOM

¹Nisrin Adelyna Darayani, ²Rini Amelia

¹English Education Department, Sebelas Maret University

²English Education Department, Sebelas Maret University
adelyna72@gmail.com

Abstract - This article is concerned with the use of feedback as one of teachers' ways in assessing students' writing in ESP classroom. This article mainly aims to investigate teachers' belief in implementing feedback for students' writing. The subject of this study is teacher in ESP class, especially Economy Sariah class. The research was conducted by using descriptive qualitative method. Interview and observation were the instruments used to gather the data. The result showed that it gave significant improvement on students' writing. Based on the class observation, it could be seen that the teacher really implemented what he believes.

Keywords: *ESP, feedback, teachers' belief*

I. INTRODUCTION

Teaching ESP students where English is not frequently taught is considered as one of the most challenging teaching practices. The students mostly have low motivation and often feel discouraged to study English. Regarding its differences from Bahasa Indonesia, the students assume that English is a difficult subject to learn and they do not have a lot of knowledge to deal with it.

As English have four skills, writing must be the last productive skill to be taught. This skill seems to be very complicated and intricate for the students. They lack of vocabulary, have limited grammar knowledge, and surely have no idea whether what they write is correct or the other way around.

Concerning the condition in ESP environment above, ESP teachers get to take the responsibility for creating the best atmosphere in ESP class. It means that the teachers get to be very careful with the techniques they use in the class, including the way the teachers assess their students' writing.

By looking at the problems exposed above, this article aims to investigate the teachers' belief on giving feedback to their students' writing in ESP class and describe what they believe in the kind of feedback they apply in the class.

In teaching and learning process, teachers' belief is one part that cannot be separated from. As stated by Calderhead (1995:77), teachers' belief refers to teachers' pedagogical belief or the belief of relevance to an individual's teaching. Additionally, Donaghue (2003:101) mentioned that teachers' belief is also about the assumption made about their students. Teachers' belief is also believed as one important thing for understanding and improving teaching learning process. They can be used as guidelines for language teachers to adopt their classrooms practices and to cope with daily language teaching problems (Fauziati: 2015).

Another definition comes from Borg (2001: 186). He defines teachers' belief as a proposition which may be consciously or unconsciously held, is evaluative in that it is accepted as true by the individual, and is therefore imbued with emotive commitment; further, it serves as a guide to thought and behavior. In accordance with Borg's idea, Schwitzgebel (2011) reveals that belief is the acceptance of something as true, or thinking that something can be true.

There are many things to deal with when teachers implement their beliefs. It is because what teachers do in the class including selecting material, preparing teaching-learning activity, assessing students' learning, and giving feedback reflects their belief (Pajares and Kagan, 1992; Borg, 2001).

Regarding some experts' views on teachers' belief, the researchers assumed that teachers' belief is a conceptual tool from teachers in making decision, behaviors in the class, and making interaction with students.

Being known as a general term, a feedback which focuses on correction is called corrective feedback (Fungula, 2013). Corrective feedback (CF) refers to teacher and peer responses to learners' erroneous second language (L2) production (Li, 2013). Written corrective feedback refers to hereafter as 'written CF' and also known as 'grammar correction' or 'error correction' (Ferris, 2012). Corrective feedback is information given to learners regarding a linguistic error they have made (Loewen, 2012; Sheen, 2007; Lightbown and Spada, 1999).

Related to the assessment of students, corrective feedback in writing classroom has a crucial role. F. Hyland & K. Hyland (2006b, p. 206) states, "by pointing forward to other texts students will write, assisting students to work out the text's potential and to comprehend the writing context, and providing a sense of audience and an understanding of the expectations of the communities they are writing for". In addition, (Hyland, 1998; Ferris, 1997) add that corrective feedback also functions as a medium of interaction between teacher-students or the contrary that can improve students' language ability. When the properties of corrective feedback have been well chosen, it contributes to the learning process (Balzer, Doherty, & O'Connor, 1989; Kluger & Denisi, 1998).

Many studies reveal that feedback has positive and negative impacts when implemented in classrooms. Sheen (2007, p. 255) proves that corrective feedback improves students' accuracy. Moreover, Seker and Ayca (2014) think that feedback is beneficial for the improvement of students' foreign language.

On the other hand, some studies show that corrective feedback does not give positive impact. Based on Al-Bakri (2015), the study revealed that there is lack of communication between teachers and students regarding WCF. Moreover, students' attitudes towards written corrective feedback can have a negative emotional impact on teachers. Although teachers generally practice what they believe, their practices sometimes do not match their beliefs when it comes to peer feedback in L2 writing and teachers' practices may change over time if their beliefs change (Shulin, 2013).

II. RESEARCH METHODOLOGY

This research was conducted to an ESP lecturer who teaches writing at the University of Muhammadiyah Malang. The researchers applied descriptive qualitative in order to present the information about a particular phenomenon (Merriam, 1998).

The data was collected through semi-structured interviews. An interview guide was used to elicit teachers' belief about feedback, experience of implementing feedback in the class, the reason for using feedback, as well as the effectiveness of feedback.

To validate the data, the classroom observation was conducted. The result was not much different from the result of the interview.

III. RESULT AND DISCUSSION

The result of this research is given along with the questions of problem.

Question: *What is the teachers' perspective regarding feedback used in ESP writing class?*

From the interview, the teacher thinks that feedback is very beneficial for her students to achieve the learning goal and to improve students' English ability. By implementing feedback in ESP writing class, students become more aware of their own writing. The teacher explained that, *"mmm...teaching ESP class is a challenge since we are not only focusing on the English teaching but also providing English in the real situation. Well, my students' ability is considered to have low in English ability. Looking at my students' English ability, in my opinion, feedback is one of the best ways to be implemented in my writing class. Moreover, it is an ESP class so, feedback is very important to overcome this problem"*.

Question: How do you implement feedback in your classroom?

In implementing the feedback, teacher not only uses written feedback to assess his students' writings but also oral feedback to sharpen the students' understanding.

"Once they have done with the assignment, I read their writing and give feedback on it. I usually give symbols, like circling the word, giving underline and question mark. It doesn't stop there. I also give them oral feedback by calling them one by one to have discussion. I do it because they still do not understand why the symbols come up on their writing".

Question: How does your belief influence your students' writing ability?

The implementation of feedback in the class shows improvement on students' writing ability. It can be seen from the teachers' statement, *"Slowly but surely, I can assume that almost all of them have a significant improvement on their writing. For example, the use of tense, it is the common error they made on their writing. They often put incorrect word on the sentence but after getting feedback, they can minimize their own error by themselves".*

In accordance to the interview result, the observation also showed positive result when students got feedback from the teacher. They were very excited on waiting of their turn to have feedback from teacher. After getting the feedback, they revised and analysed their error. Some of them were also having small discussion with their friend toward the feedback and error they made.

IV. CONCLUSION

This study has explored and elaborated EFL teachers' beliefs and practices with regard to the implementation of feedback activities in one ESP writing class. The teacher in this study has certain belief that oral and written feedback he usually implements can give significant improvement for his students' writing. Based on the class observation, it can be seen that the teacher really implemented what he believes. And meeting by meeting, students could be more aware of either the mistakes or errors they frequently made. This results in their better achievement of writing.

REFERENCES

- [1] Ferris, D., "The influence of teacher commentary on student revision," TESOL Quarterly, vol. 31, pp. 315-339, 1997
- [2] Ferris, D., "Written corrective feedback in second language acquisition and writing studies," Language Teaching. Cambridge University Press, vol. 04, pp. 446-459, 2012.
- [3] F. Hyland, and K. Hyland, "Feedback on second language students' writing," Language Teaching. vol. 39, pp.89-101, 2006.
- [4] Fauziati. Endang, Teaching english as a foreign language: principle and practice. Surakarta: Era Pustaka, 2015.
- [5] Meriam, B. Sharan, Qualitative Research and Case Study Application in Education. San Francisco: The Jossey-Bass, 1998.
- [6] Li. Shaofeng, "Oral corrective feedback," ELT Journal Advance, Oxford University Press, pp. 1-3, 2013.
- [7] Calderhead. J, "Teachers: Beliefs and knowledge," In D. C. Berliner, & R. C. Calfee (Eds.), Handbook of educational psychology pp.709-725. New York: Simon & Schuster Macmillan, pp. 709-725, 1996.
- [8] Borg. M, "Key concepts in ELT teachers' beliefs," ELT Journals, vol. 55(2), pp. 186-188, 2001.
- [9] Donaghue. H, "An instrument to elicit teacher's beliefs and assumptions," ELT Journals, vol. 57(4), pp. 344-351, 2003.
- [10] Sheen. Y, "Corrective feedback and learner uptake in communicative classrooms across instructional
- [11] Settings," Language Teaching Research, vol.8(3), pp. 263-300, 2007.
- [12] Schwitzgebel. E, "Belief," The Stanford 2003, Encyclopedia of Philosophy: <http://plato.stanford.edu/archives/win2011/entries/belief/>
- [13] Pajares. M. F, "Teachers' beliefs and educational research: cleaning up a messy construct," Review of Educational Research, vol. 62(3), pp. 307-32, 1992.
- [14] Fungula. B.N, "Oral Corrective Feedback in the Chinese EFL Classroom," Retrieved from <http://www.diva-portal.org/smash/get/diva2:693017/FULLTEXT01.pdf>, 2013.
- [15] Kagan. D. M, "Implications of Research on Teacher Belief," Educational Psychologist. Vol. 27, pp. 65-90, 1992.
- [16] Lightbown, P. M., & Spada, N, How languages are learned. Oxford, UK: Oxford University Press, 1999.
- [17] Loewen. S., and Philp. J, "Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness," Modern Language Journal, vol. 90, pp. 536-555, 2012. Retrieved from <http://dx.doi.org/10.1111/j.1540-4781.2006.00465.x>
- [18] Kluger, Avraham N., and Angelo DeNisi, "The Effects of Feedback Intervention on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory," Psychological Bulletin, vol. 119 (2), pp. 254-84, 1998.

- [19] Balzer, Doherty, & O'Connor, "Effects of cognitive feedback on performance," *Psychological Bulletin*, Vol. 106(3), pp. 410-433, 1989.
- [20] Seker. Meral and Ayca Dincer, "An insight to students' perceptions on teacher feedback in second language writing classes," *English Language Teaching, Turkey: Canadian Center of Science and Education*, vol. 7(2), pp. 73-82, 2014.
- [21] Hyland, K., & Hyland, F, "Interpersonal aspects of response: Constructing and interpreting teacherwritten feedback." In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues*, pp. 206-224, 2006. Cambridge, UK: Cambridge University Press.
- [22] Shulin. Yu, and Icy Lee, "Understanding EFL students' participation in group peer feedback of L2 writing: A case study from an activity theory perspective," *Language Teaching Research*, Vol. 19(5), pp. 572-593, 2013.

Comparison of Character Value Between Lower Class And Upper Class At Salman Al Farisi 2 Elementary Integrated School

Rosaria Irjanti¹, Farida Agus Setiawati²
Magister Psychology at University of Yogyakarta
irjanti@gmail.com¹, farida_as@uny.ac.id²

Abstrac- The research was backed by the Law of the Republic of Indonesia Number 20 Year 2003 on National Education objectives of the formation of character. This study was photographing a character value honest, intelligent, independent and decent-looking in learners in SDIT Salman Al Farisi 2, Yogyakarta. This study using 38 samples on learners in grade 1 to grade 6. The research approach uses quantitative methods to determine how high the character values that are embedded in the learners in SDIT Salman Al Farisi 2. Model t-test to analyze the differences in the value of character lower class and upper class. This study shows the value of the dominant character in the character polite than honest, autonomy and intelligent. The study also earn the difference between the value of the general character with the data analysis on the score t-test ($t = 2.770$; $p = 0.009$) where $p < 0.05$ which indicates that there is a difference between the value of lower-class characters with upper class.

I. INTRODUCTION

Character of the nation is an important pillar in the life of the nation. The nation's character will form citizens who are creative and innovation that benefit the nation. The importance of character for the progress of a country because character has an important role for developed countries (Lickona (2007)).

The Indonesian government has formulated a policy values imbued national character of Pancasila and the configuration of psychological and socio-cultural (Ministry of National Education, 2010). The values of these characters is set to 18-character value. Meanwhile, according to Marzuki (2010), there are 26-character value that must be achieved in learning at school.

The character value to be achieved by each school but the implementation could be experiencing a lot of difference. This difference occurs because each institution has a priority in the implementation of character values (Ministry of National Education, 2011)

Implementation of character education in schools based on educational objectives contained in Government Regulations No. 20 of 2003. The provision of appropriate learning with educational containing characters and pay attention stages of moral development of students can form characters expected and eventually to help learners resolve their problem.

Moral development of a person affected by the intellectual growth, peer interaction and curtailing the power of adults. On the theory of moral development Piaget is known that increasing the age of the child then also increased the understanding of the rule. At first, the child is at the level of blind obedience then their knowledge of his moral growing, and to understand that legislation may be compromised (Dhuskan and Whelan, 1984; Berk in Mc Innerny, 2006).

The theory of moral development according to Kohlberg also shows the person's moral development at three levels which are divided into six stages. Prekonvensional level consists of heteronomous morality and individualism. The conventional level consists of the third and fourth stages are stages expectations and mutual expectations. Pascakonvensional level consists of stages five and six of the stages of the contract and universal ethical principles (Kohlberg and Hersh, 2007).

The first stage occurs at an early age until the end of the age of the child is age 2 years to 13 years. In the first stage something right or wrong based on social values around and see the

cause and effect of the impact of such conduct as punishment. The second stage is the stage where the children understand that when they do good deeds then they get a good thing, if they do something bad then they get a bad thing. Third and fourth stages occur in the early teens to late teens, the age group 13 years to 20 years. In the fifth and sixth stage is a stage that requires an understanding of the moral high enough to understand an action is performed (Dhuskan and Whelan, 1984; Santrock, 2011).

Based on theory of moral development raised by Piaget and Kohlberg, it is known that age is one aspect in the stage of moral development. At an early age, children are more obedient to the rules but with age is accompanied by social interaction, the understanding of a compliance acts based on predefined rules evolved that an act based on social values that can be compromised.

In addition to aging, the formation of character in children so that it becomes personality is influenced also by the environment. If the child lives in a social environment that is characterized, it will form a child's character. Therefore, it needs awareness of all parties that affect the child's life such as family, school and entire community or institutions to shape the character of good (Zuhdi, Prasetya, and Masruri, 2012). This is consistent with ecological theory on mikrosistem and mesosistem (Bronfenbrenner, 1994).

Students who have good character will be able to improve academic skills (Benninga; Jacques, S; Berkowitz, M.W; Phyllis; Smith, 2003) and the ability of soft skills (Mutakin, 2014). Not only academic skills as well as soft skill ability but someone who has a good character will have a positive mental attitude. Positive mental healthy will provide benefits not only to himself but also to the surrounding environment (Ryan and Deci, 2006).

Therefore, it can be said the character is a character that contains the values that form of social values in accordance with the stage of development of a person's life. Students who have good character will benefit both the academic, soft skills and a positive mental attitude to himself and the surrounding environment.

SDIT Salman Al Farisi 2 is one of the agencies under the auspices of the Foundation Salman Al Farisi that emphasizes character education as a priority in the learning process. Character Honest, Intelligent, Self, and Santun (later called JUSMANTUN) a four-character value that has been the watchword in all the institutions that are the foundation Salman Al Farisi. The importance of the implementation of character education is also seen from the results of student learning or raport report.

SDIT Salman Al Farisi 2 as a school that has made the character as a hidden curriculum. Therefore, it is essential to see how far the implementation of the character values in students in SDIT Salman Al Farisi 2, this needs to be done in connection with the report of educators on the teachers forum both upper class and the lower class, that still there are some children who do not behave prosocial acts such as saying no polite and do dishonesty.

Based on the background of the problems mentioned above, the researchers are interested in examining how high the value of the student's character based on the character that has been proclaimed by the Foundation Salman Al Farisi that JUSMANTUN in the learning process in SDIT Salman Al Farisi 2. Therefore, this study aims to see character values honest, intelligent, autonomy, and polite on the lower classes (grades 1-3) and upper classes (grades 4-6) and compare the values of characters in the upper class and the lower class in SDIT Salman Al Farisi 2.

II. RESEARCH METHODS

The research approach using quantitative methods. It is research using numeric data to analyze the conditions based on the observations that have been made. Quantitative methods to determine how high the character values that are embedded in the learners in SDIT Salman Al Farisi 2 and comparing the upper and lower classes.

This research was conducted in SDIT Salman Al Farisi 2, Wedomartani, Jetis, Ngemplak, Sleman, Yogyakarta. This school has been taken because of the character education program, it is seen from student's raport card. In the report card students are academic value and the value of the personality of students

Data were collected since February and May 2015. The study was conducted in the second half. This is done because the observer has to know the subject for more one semester.

Observer is an educator who has known the students at minimum six months during the period teachings.

The subjects in this study were students in SDIT Salman Al Farisi 2 from grade 1 to grade 6. While the technique of data collection was done by using observations made by educators. Educator observation is homeroom students who have understood the conditions subyek more than six months.

The research instrument is the observation guidelines that form the character scale using Likert-type. The response on the observation of this scale in the form: Always (S), Frequently (SR), Rare (J), and Never (TP). Before the instrument is used, it should test the validity and reliability.

The validity of the instruments carried by content. Validitas of the content in this study using content validity coefficient by Aiken. The Analysis obtained an average of validity index was 0.81. Some item that has a low coefficient has been revised and carried back to the expert assessment.

The reliability test is intended to measure the degree of constancy of the measuring instrument. the results of the reliability test showed a measure may provide an alternative no different when performed repeated measurements on the same subject (Anwar, 2004). Testing reliability of this scale using Cronbach Alpha. Through this technique, reliability for variable item value character is 0,877. Based on the analysis of reliability show that high reliability. Therefore, that the instrument is steady and trusty.

III. CONCLUSION

After all the research data collected, the data scores and tabulated. After that, the data were analyzed to see the characters appear. Analysis of the data in this study are using descriptive and independent sample t-test.

A. Descriptive research subjects

Descriptive research subjects that provide a picture of the state of research. Subjek subject of this research were 38 subjects. Subjects were aged between 7-12 years. Consisting of 21 samples of the lower classes and 18 upper classes .The picture of the subject of the study are presented in the following table:

Table 1. Descriptive research object on the basis of the level of the class

Study Class	Subyek
	Amount (%)
Lower class	52,63 %
Upper class	47,36 %
Amount	100%

In table 1 shows the percentage of lower class number 52, 63 percent of 100 percent, while the upper class numbered 47,36 out of 100 percent of the number of subjects. From Table 1 shows that the subject's lower class more than the upper class.

B. Description of statistics

Table 2. Group Statistics

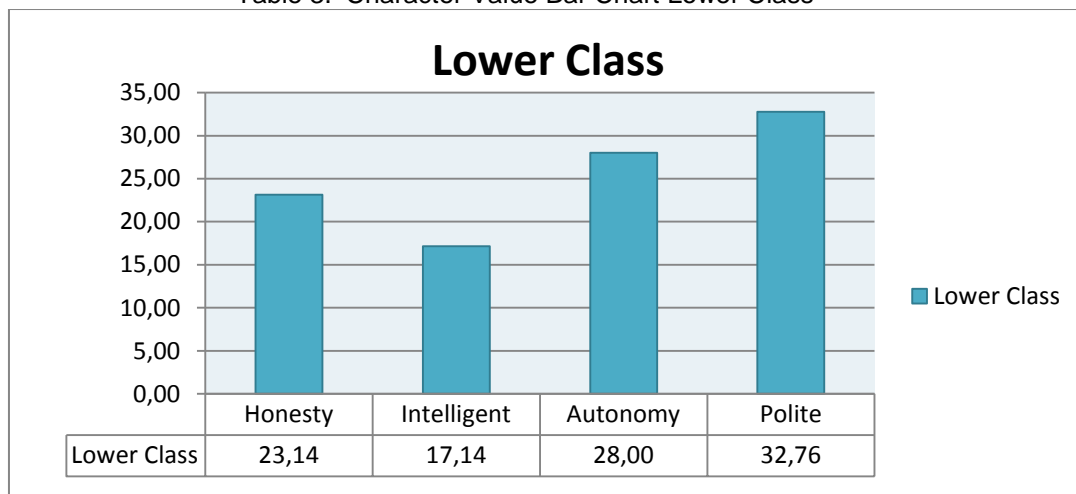
Study Class		N	Mean	Std. Deviation	Std. Error Mean
Honesty	Lower Class	21	23,1429	4,25777	,92912
	Upper Class	18	20,3333	3,95564	,93235
Intelligent	Lower Class	21	17,1429	4,39643	,95938
	Upper Class	18	14,3889	5,38122	1,26837
Autonomy	Lower Class	21	28,0000	4,75395	1,03740
	Upper Class	18	23,0000	6,29659	1,48412
Polite	Lower Class	21	32,7619	4,94879	1,07991
	Upper Class	18	28,1111	6,60560	1,55696
The general character	Lower Class	21	101,0476	15,33452	3,34627
	Upper Class	18	85,8333	18,97134	4,47159

Table 2 is a description statistical group table which shows the mean value of the character of each aspect and general character. On the value of honest character, mean the lower classes, namely 23, 1429 while the mean upscale, namely 20, 3333. Mean value of intelligent character in the upscale namely 14, 3889 while the lower class mean, ie 17.1429. Mean self-contained character in the lower class of 28, 0000, while the upper class mean by 23, 0000. On mannered character, the lower class has a mean 32, 76 919, while the upper class had a mean of 28.1111.

In the mean general character of the lower class looks higher at 101, 0476 compared with the upscale ie standard deviation is 28.1111 15.33452 6.60560 for the lower classes and for the upper class.

Based on statistical group in Table 2 can be prepared bar charts on the value of the character that appears in the lower class and upper class.

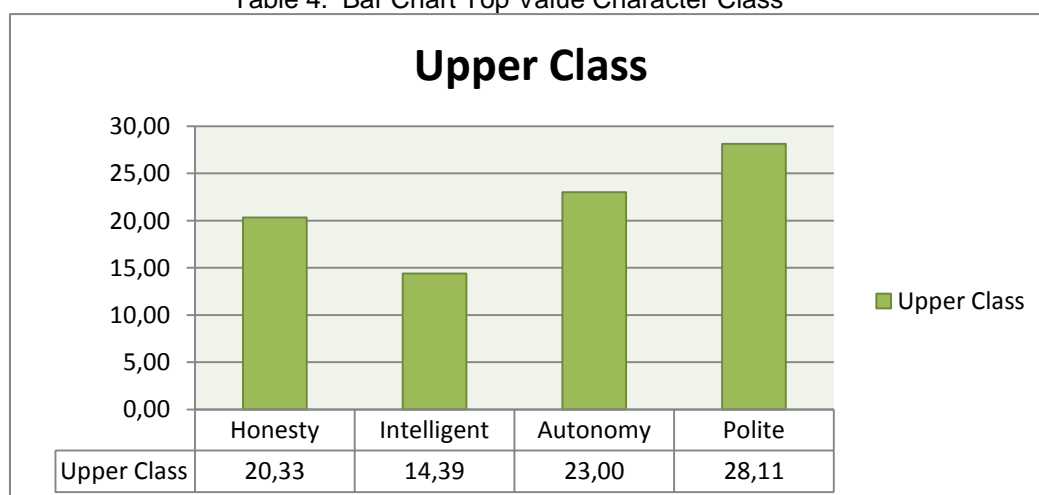
Table 3. Character Value Bar Chart Lower Class



In Table 3 shows that the character Santun has a mean value of the highest of 32.76. Under politeness value was the value of autonomy, 28.00. The value of honesty has a mean value of 23, 14, it was higher than the mean intelligence, 17, 14.

The results of observations conducted by educators in the upper class in SDIT Salman Al Farisi 2, can know the value of a character that appears in the lower class in the form of a bar chart:

Table 4. Bar Chart Top Value Character Class



In Table 4 shows that the value of character polite has a mean value of the highest of 28.11. Under politeness value was the value of autonomy, 23.00. The value of honesty has a mean value of 20.33, it is higher than the mean intelligence is 14.39.

Based on table 2, 3 and 4 are known that the character who seemed to stand on the lower classes and upper classes is value polite character. Polite character value indicates lower class and upper class was 32.7619 and 23.0000. This condition does not correspond to the alleged educators teachers forum that students in SDIT Salman Al Farisi 2 has a value lower than the polite character with another character value. In addition, the difference in value between the lower-class characters and upper class. Value character at lower-class was higher tendency than upper class.

C. Description Based On Criteria

Based on the calculations have been done research and data obtained from the questionnaire study, obtained a description based on criteria. In this study, using the criteria for categorizing average. This categorization is to put individuals into groups whose position is tiered according to a continuum based on the attribute being measured (Anwar, 2004). The categorization used in this study refers to three categories: low, medium, and high. General overview of the study can be seen in the table descriptions of research data includes the value of honest character, intelligent, independent, and autonomy on the lower classes and upper class. The are presented in the following table:

1. Description Character Aspect

Based on the analysis using the criteria, known to the average general character was 94. Based on these criteria, then drafted three categories: low, medium, and high. To lower the value of 78, the categorization of being with a range between 78 to 110, while the categorization is high, that is above 110.

Table 5. Based Categorization Criteria

	Criteria	Categorization	Amount Of Lower Class	Amount of Upper Class
Character	$x < 78$	Low	2	6
	$78 > x > 110$	Middle	10	9
	$x > 110$	High	8	3

Table Shows that the value of the general character in students at SDIT Salman Al Farisi 2 dominant was the categorization of being, that were the lower class numbered 10 and upper class totaled 9. In categorization lower in the upper classes for the characters in general tend to be higher than the lower classes. While the value of the character in general to high categorization shows that the lower classes that tend to be more dominant than upper classes. This shows that students in lower classes more character than the students in the upper classes.

D. Description Each Aspect

Description of every aspect made based categorization using the criteria in every aspect. Based on an analysis using the categorization criteria for low, medium and high for each aspect. This categorization is seen in the lower class and upper class. The average value of the aspect truthful lower class was 23, and upper class was 20. In the aspect of intelligent lower classes was 17 and upper class was 14. Aspects of autonomy at lower classes are 28 and upper classes was 23. Aspects of polite at lower class was 32 and the upper class was 28.

Table 6. Categorization Each Aspect

Character	Categorization		Lower Class	Upper Class
Honest	$x < 18$	Low	4	7
	$18 > x > 25$	Middle	6	7
	$x > 25$	High	10	4
Intelligent	$x < 12,5$	Low	4	6
	$12,5 > x > 19,5$	Middle	8	8
	$x > 19,5$	High	8	4
Autonomy	$x < 23$	Low	3	9
	$23 > x > 29$	Middle	9	5
	$x > 29$	High	8	4
Polite	$x < 29$	Low	3	9
	$29 > x > 38$	Middle	12	7
	$x > 38$	High	5	2

Table 5 shows the categorization of every aspect based on the average. In every aspect of honest, intelligent, autonomy, and polite known that high categorization dominated more lower class than the upper class. Categorization high in the lower class that is numbered 10 aspects honest, intelligent aspects were 8, amounting to 8 autonomy aspects, and aspects polite totaled 5. High above categorization in class that is consists of 4 aspects honest, intelligent aspect numbered 4, independent aspect consists of 4, while the manners aspect numbered 2.

Based on Table 5, it can be concluded that the lower class is higher than the total value of upper classes character. This is evident from the large number of high aspect categorization honest, intelligent, autonomy and polite.

E. Comparison Values Character Students at the Upper Class and the Lower Class

independent sample T-test used to know difference between the character values lower class and upper class. T-test using a significance is 0.05. Based on the test independent sample t-test, with significance value is 0.05, it is known that the value of t are 2.770 and p value is 0.009, where $p < 0.05$ which indicates that there are differences between the mean value of the overall character of the lower class to upper class.

The results are consistent with the theory of moral development according to Piaget and Kohlberg someone that moral development of a person experiencing various stages of development (Dhuskan and Whelan, 1984; Berk in Mc Innery 2006; Kohlberg and Hersh, 2007). The different stages of this development is caused due to aging. Therefore, in the stage of moral development according to Piaget and Kohlberg that the difference between the lower class and upper class might happen. Such differences are students of lower classes more character than the students in the upper classes.

Based on table 2, 3 and 4 is known that the character who seemed to stand on the lower classes and upper classes is value mannered character. Polite character value indicates lower class and upper class was 32.7619 and 23.0000. This condition does not correspond to the alleged teachers in teacher forum that learners in SDIT Salman Al Farisi 2 has a value lower than the mannered character with another character value. In addition, the difference in value between the lower-class characters and upper classes. value character at lower class is higher tendency than upper classes.

IV. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

SDIT Salman Al Farisi 2 has 320 students. SDIT Salman Al Farisi 2 is one school institutions under the Foundation Salman Al Farisi, has endorsed the importance of education karakter. Character education is not only listed in vision and mission Foundation Salman Al Farisi of "Character and Achievement". Preferred character is the character of an honest, intelligent, autonomy and polite but also present in the learning outcomes of students provided by the school end of each semester.

Based on the analysis that has been done, the researchers concluded that among aspects of the character of an honest, intelligent, autonomy and polite, a prominent character in the students in SDIT Salman Al Farisi 2 is a polite character.

This research method using quantitative methods. Analysis of descriptive statistics and t-test to see which characters appear and the difference in value characters between the lower-class and upper classes. This study used 38 samples of the students in grade 1 to grade 6. The characters that appear in students can be seen from the calculation observations made by educators.

Descriptive statistical results showed that the distinct character of the character JUSMANTUN was characters polite. This shows that the characters polite has been possessed by students in both the lower classes and upper classes. As intelligent character is a character that needs to be improved in both the lower classes and upper classes. In the t-test analysis shows there was a difference between the character values lower class and upper class. Lower-class character value is higher than the value of upper-class character.

B. Recommendation

Based on the results of research conducted on some suggestions that the authors pose to schools, educators, and subsequent research, namely:

1. For schools that every child has the stage of moral development, therefore it is important for schools to form a culture of education to enable the student to motivate him to get better.
2. For educators, the character will be formed through habituation and moral development of children is based on a sample of adults in this case educators and parents, then giving a good example to be ongoing. Meaningful learning activities should be emphasized so that learning materials are easy to learn.
3. For further research, you should pay attention to the limitations of this research are:
 - a. The study only took a sample of 32 samples. 32 of these samples compared with the number of students in SDIT Salman Al Farisi 2 as much as 320, has not significantly photographed overall character evinced by students in SDIT Salman Al Farisi 2.
 - b. The study only looked at four characters based on the slogan launched by the Foundation Salman Al Farisi. The implementation of the research should not only see the four characters that is honest, intelligent, autonomy and polite, because SDIT Salman Al Farisi 2 have students study reports in the form of morals or manners given by the school to the parents / guardians of students. The report covers the value of the character that is religious, confident, responsible, respect, courtesy, competitive, clean and honest.

DAFTAR PUSTAKA

- [1] Benninga, Jacques S; Berkowitz, Marvin W; Kuehn, Phyllis; Smith, Karen. The Relationship Of Character Education Implementation And Academic Achievement. *Journal of Research in Character Education*; 2003; 1, 1; ProQuest Education Journals.
- [2] Brofenbrenner, U. 1994. *Ecological Models of Human Development*. International Encyclopedia of Education, Vol.3, 2nd. Ed. Oxford: Elsevier
- [3] Bruning, Schraw, dan Norby, 2011. *Cognitive Psychology and Instruction*. Pearson: Boston.
- [4] Dhuskan dan Whelan, 1984. *Moral Development. A Guide to Piaget dan Kohlberg*. Kanisius: Yogyakarta.
- [5] Kementerian Pendidikan Nasional. 2011. *Panduan Pelaksanaan Pendidikan Karakter*. Badan Penelitian dan Pengembangan Pusat Kurikulum dan Perbukuan. Jakarta.
- [6] Kementerian Pendidikan Nasional. 2010. *Kerangka Acuan Pendidikan Karakter*. Jakarta.
- [7] Mc Inerney, Dennis. 2006. *Development Psychology For Teachers: An Applied Approach*. Allen&Uwin: Singapore.
- [8] Mutaqin. 2014. Implementasi Pendidikan Karakter Dalam Pembelajaran Berbasis Pratik Untuk Meningkatkan *Soft Skill* Mahasiswa. *Jurnal Pendidikan Karakter*, Tahun IV, Nomor 2: Universitas Negeri Yogyakarta.
- [9] Ryan, Richard M. and Deci, Edward L. 2006. Self-Regulation and the Problem of Human Autonomy: Does Psychology Need Choice, Self-Determination, and Will?. *Journal of Personality* 74:6. Blackwell Publishing, Inc.
- [10] Saifudin Azwar. 2014. *Penyusunan Skala Psikologis*. Yogyakarta: Pustaka Pelajar.
- [11] Thomas Lickona. 1993. The Return Of Character Education. *Character Education*. Volume 51, Number 3, Pages: 6-11.
- [12] Santrock, John W. 2011. *Life Span Development*. Jakarta: Penerbit Erlangga. Inc.

Authentic Assessment in the Learning of Social Studies

Rudy Gunawan

Magister Program of Social Studies UHAMKA, Jakarta
rudyansich@gmail.com

Abstract - The purpose of this research is to know how authentic assessment given in social studies. Learning social studies in junior high school includes materials of History, Geography, Economics and Sociology. In mono-disciplinary, each of these materials have characteristics and a different way in the planning, implementation and evaluation. But in term of interdisciplinary, evaluation was made in order to provide a thorough assessment. Assessments can be done in schools is an authentic assessment. Authentic assessment is often also referred to as performance-based assessment. In an authentic assessment of students are required to show meaningful tasks then assessed directly in the classroom. Nevertheless, authentic assessment is not only implemented in the classroom, but also can assess the tasks performed outside the classroom with project-based learning. The method used in this research is a survey conducted for teachers of social studies in Jakarta. The results showed that an authentic assessment can help students to display meaningful tasks then assessed directly in the classroom. However, authentic assessment is not only implemented in the classroom, but also can assess the tasks performed outside the classroom with project-based learning.

Keywords: *authentic assessment, learning, social studies*

I. INTRODUCTION

Curriculum 2013 is designed to provide reinforcement to the students so that experiences a competency attitudes, knowledge and skills. The concept of learning in elementary school and junior high school is different in the form of presentation. At primary school level, all subjects are combined and presented as a theme. Junior high school level, presenting to students is different by separating the material into a single course. But this separation has not been done completely, there are several related fields are combined into a single entity such as science and social subjects which is a combination of several subjects.

Social studies examine a set of events, facts, concepts, and generalizations relating to social issues. In junior high school social studies contain materials Geography, History, Economics, Anthropology and Sociology with integrated learning. Social studies designed to develop knowledge, understanding and capability analysis of social conditions. The capabilities required to enter the life of a dynamic society. Social studies, arranged in a systematic, comprehensive, and integrated in the learning process towards maturity and success in life in society. With this approach the expected learners will gain a broader understanding and depth in science-related fields (Lampiran 3 Permendiknas No 22 Tahun 2006 tentang Standar Isi, 2006).

This is consistent with the vision of education as a social studies education program that focuses on the development of the individual as a social actor capable of taking reasoned decisions and as intelligent citizens who are committed, accountable and participatory (Winataputera, 2009). The breadth and scope of the social studies material, causing some of the obstacles encountered in learning social studies teacher. In a study conducted by Pratiwi (2012) teachers difficulties when preparing lesson plans as much material as well as difficulties in presenting the material to divide their time. This causes the students did not achieve the expected competencies.

Based on preliminary observations made on May 5, 2016 to several social studies teachers are currently studying in the Master of Education Program for Social Studies UHAMKA, many elements must be assessed starting from attitudes to skills in one meeting. Teachers still have not found an effective way to implement an authentic assessment. This is in line with the results of research conducted by Kusmijati (2014) had the most difficulty many complaints by the

teachers is the understanding of the core competencies and basic competency. Master difficult how to teach it and make an assessment. Additionally, detailed guidance is not owned by the teacher. In another study, teachers had difficulty in implementing authentic assessment because of the low creativity of teachers, characteristics of students who do not support, the lack of training of authentic assessment and insufficient time (Enggarwati, 2015).

Nevertheless, authentic assessment effectively to assess learning social studies. According to a survey conducted by Widoyoko (2007) with respondents social studies teacher as much as 6 (six) indicates that an evaluation of the quality of learning social studies less attention. social studies assessment of learning outcomes is more focused on aspects of academic skills (knowledge) and less attention to other skills. Based on this it is necessary in an authentic assessment of learning social studies so that an assessment not only to the competence of knowledge but also includes attitudes and skills competency.

Based on the background described, in this study the formulation of the problem posed is:

1. How does the knowledge of authentic assessment?
2. How is the planning, assessment, advantages and constraints of teachers authentic assessment done?
3. How do the authentic assessment of learning social studies?

The purpose of this study was to describe an authentic assessment of learning is to know social studies:

1. Knowledge of authentic assessment
2. Planning, assessment, advantages and constraints of teachers authentic assessment conducted
3. Assessment of authentic learning social studies

The results of this study are expected to provide academic and practical significance. In academic research findings are expected to increase the knowledge and analytical sharpness associated with the problem, especially in an authentic assessment of learning social studies. In practical terms this study are expected to be input for teachers social studies on the implementation of authentic assessment.

II. RESEARCH METHOD

The method used in this research is descriptive survey research method that aims to provide a picture or a description of a situation (phenomenon) objectively and identify problems to get justification situation and the ongoing practice. The population object of the study was teachers who are studying social studies S2 in the Master Program of Social Studies Education UHAMKA. Sampling using simple random sampling technique for members of a relatively homogeneous population so it can be taken at random without regard to strata that exist in the population. The sample used in this study is 18 samples out of 35 total population.

The instrument used to measure the function of the teacher's knowledge of authentic assessment. In this study the data and information collected from respondents using questionnaires distributed via email with the help of google form. The questionnaire contains a list of questions that assess teachers' knowledge and experience in conducting assessments social studies authentic. The type of questionnaire used a combination of closed and open questionnaire. Material questionnaire tailored to the theory of authentic assessment. Data processing techniques with descriptive analysis that aims to describe the data that has been collected without intending to generally accepted conclusion.

III. RESEARCH FINDINGS AND DISCUSSIONS

According to the survey, respondents who filled out a questionnaire about 18 people consisting of 13 people (72.2%) were male and 5 (27.8%) women. Works 100% of respondents as a teacher (a teacher or lecturer) in various levels of education and all respondents social studies teaching or social sciences (economics, geography and history). Experience working as a teacher a majority of less than 5 years as many as eight people (44.4%), between 5-10 years as many as three people (16.7%), between 10-15 years of as much as 2 people (11.1%) and in over 15 years as many as five people (27.8%). All respondents are already using authentic assessment in the classroom with the category of "always" 5 people (27.8%), "often" as many as 10 people (55.6%) and "sometimes" as many as three people (16.7%). Briefly condition of respondents can be seen in the following table:

Table 1. Description of Respondents

No	Description	Options Answers	Total (person)	Percentage (%)
1	Gender	Man	13	72,2
		Female	5	27,8
2	Work	Primary school teachers	3	16,7
		Junior high school teachers	4	22,2
		High School teachers	6	38,9
		Lecturer	1	5,6
		other educators	3	16,7
3	Length of work	Less than 5 years	8	44,4
		5 to 10 years	3	16,7
		10 to 15 years	2	11,1
		Above 15 years	5	27,8
4	The use of Authentic Assessment	Always	5	27,8
		Often	10	55,6
		Sometimes	3	16,7
		Never	-	-

Source: Research data in 2016

Based on these data, it can be concluded all respondents already implementing authentic assessment, so it can be assumed that knowledge, planning, processes, benefits and challenges of authentic assessment is already known by the respondents.

A. Knowledge of Authentic Assessment

Questions posed to respondents with respect to the initial knowledge of respondents about the difference test, measurement, assessment and evaluation, understanding authentic assessment and authentic assessment techniques for competency attitude (spiritual and social), competency knowledge and competency skills. Respondents were given statements that contain answers to questions about the test, measurement, assessment and evaluation.

In summary the results of the respondents described as follows:

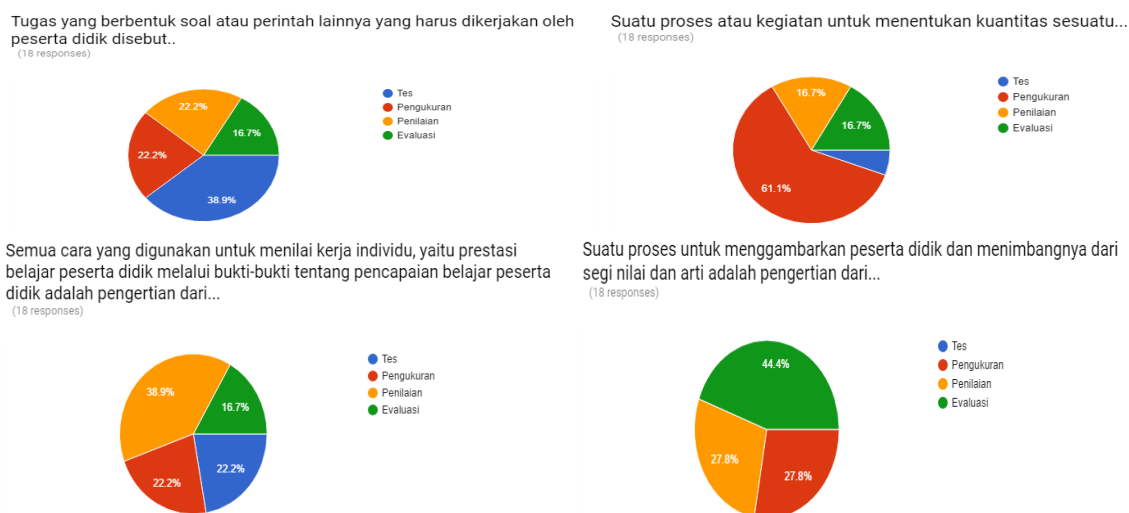


Figure 1. Knowledge of Test, Measurement, Assessment and Evaluation

The survey showed that there is still confusion of respondents to distinguish understanding of test, measurement, assessment and evaluation. Respondents who answered correctly to questions about: understanding the test as many as 7 people (38.9%), understanding the measurement 11 (61.1%), understanding the assessment of 7 people (38.9%) and understanding evaluation 8 people (44.4%).

Based on the above description, the majority of respondents did not know the difference between a test, measurement, assessment and evaluation. This is consistent with the view of Arifin (2009) that often, the term evaluation and assessment considered the same as test and measurement. Some sense of the term test, measurement, assessment and evaluation according to Arifin (2009) are as follows:

1. Said Hamid Hasan (1988) explains that the test is a data collection tool specially designed and can be seen from the construction items. So the test is a data collection tool that can berupakan questions and designed through a set of strict criteria.

2. Measurement of a number of the procedures for granting or variable attributes of a continuum (Azwar, 2012).
3. Ratings are all means used to assess individual work, namely the achievement of learners through the evidence of the learning achievements of learners (Mardapi, 2008).
4. Lincoln and Guba (1985) states that the evaluation is a process to describe learners and weighed in terms of value and meaning of evaluation is a systematic and ongoing process to determine the quality (value and meaning) of something based on certain considerations and criteria in order decision-making.

However, all respondents (100%) agree that an authentic assessment together with an assessment of performance. Authentic assessment is often referred to as the performance appraisal where an assessment is said to be authentic if it can directly observe the behavior of learners and the performance appraisal process in real situations. The performance assessment is expected to measure seven basic capabilities by Howard Gardner is visual-spatial, bodily-kinesthetic, musical-rhythmical, intrapersonal, logical mathematic and verbal linguistic (Zainul, 2001).

Knowledge about assessment techniques for all competencies that must be measured quite adequate. Respondents were asked to select more than one technique in the assessment questionnaire given. The survey results regarding authentic assessment techniques for competency spiritual and social attitudes can be seen in the chart below:

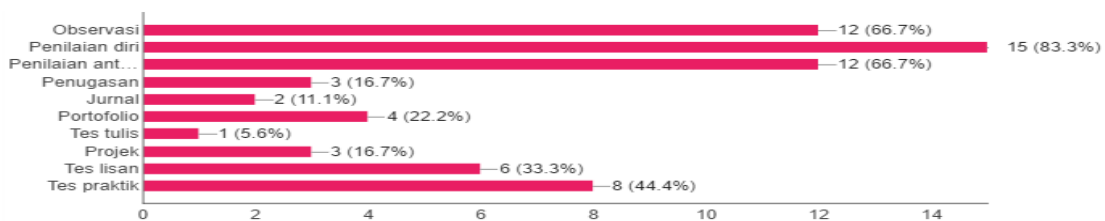


Figure 2. Survey Techniques for Authentic Assessment of Competence Attitude

The survey results regarding authentic assessment techniques for knowledge competencies can be seen in the chart below:

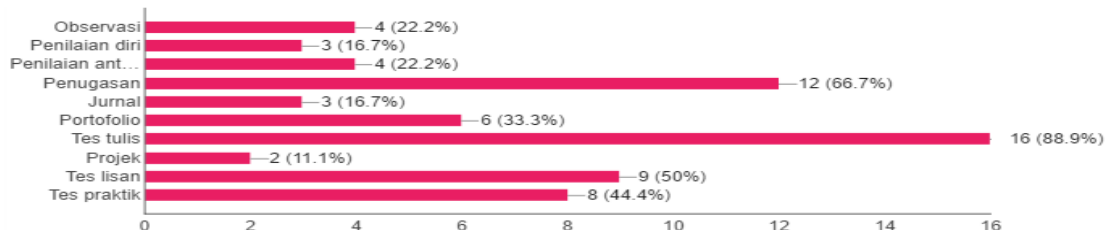


Figure 3. Survey Techniques for Authentic Assessment Competence Knowledge

The survey results show, the majority of respondents chose a written test (88.9%) and assignments (66.7%) and oral test (50%) can assess the competence of knowledge. This is in line with the criteria established by the Ministry of Education and Culture that educators assess the competence of knowledge through a written test in the form of multiple choice, stuffing, short answer, true-false, matching, and the description that accompanied the scoring guidelines. Oral test instrument can be a list of questions and instruments may include homework assignments and/or project (Lampiran Permendikbud RI No.66, 2013).

The survey results regarding authentic assessment techniques for competency skills can be seen in the chart below:

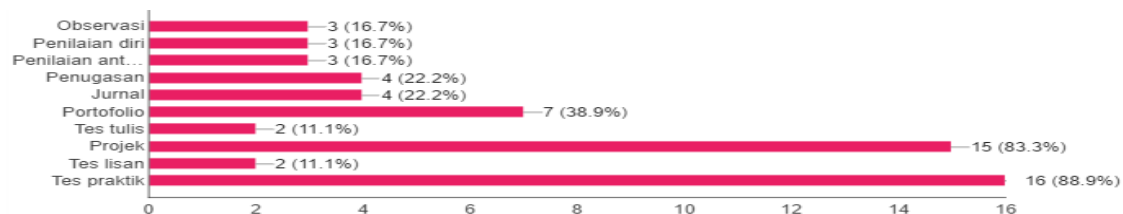


Figure 4. Survey Techniques for Authentic Assessment Competency Skills

The survey results show, the majority of respondents chose practice tests (88.9%) and projects (83.3%) can assess the competence of the student's skills. Just as much as 38.9% of respondents who choose the portfolio, portfolio assessment is an assessment that can assess the entire collection of student work that is reflective-integrative to know the interests, growth, achievement, and creativity of learners within a certain time (Lampiran Permendikbud RI No.66, 2013).

B. Planning, Process Excellence and the Barriers in Authentic Assessment

Descriptive questions given to respondents regarding the preparation of the respondents in the assessment of authenticity. Questions are open and clearly answered by the respondents. The majority of respondents replied that the planning is done merely preparing assessment instruments. Only one person (5%) of respondents who provide detailed planning. The implementation process was answered by the respondent merely provide an assessment in the classroom in the form of questions and the students do directly. The majority answered that the assessment carried out at the end of the lesson, and only 3 (16.7%) of respondents who answered that the assessment done throughout the learning process.

In fact, authentic assessment begins with the design of the cover input ratings (early learning) process (for learning) and output (after learning) (Kunandar, 2013). Teachers must know the purpose of learning and make a pre-test questions, exercises, discussion of observation rubrics, homework, worksheets and formative assessment. Authentic assessment is not limited in the classroom. In authentic learning, learners are asked to collect information with a scientific approach, memahami various phenomena or symptoms and their relationship to one another deeply, and relate what is learned to the real world outside of school. Teachers and learners have a responsibility for what happened. Learners also know what they want to learn, have a flexible time parameters, and is responsible for staying on task. Authentic assessment encourages learners to construct, organize, analyze, synthesize, interpret, explain and evaluate information to turn it into new knowledge (Kunandar, 2013).

The advantages of authentic assessment according to the results of the survey are (2016):

1. Assessment can be done thoroughly and touches all aspects of cognitive, affective and psychomotor.
2. The learning result is more accurate and original
3. Students can determine the ability of self in the learning process and the teacher can determine the learning strategies and evaluate the learning process that has been done so as to improve the quality of learning.
4. Students play an active role in the assessment process. In this phase can reduce anxiety, fear of getting bad grades that can mengganggu pride.
5. Authentic assessment successfully used by students from different cultural backgrounds, learning styles and academic ability.
6. Tasks that used in the valuation more attractive and reflect the authentic daily life of students.
7. A more positive attitude toward school and learning can flourish.
8. Teachers hold a larger role in the assessment process other than through traditional testing program. This engagement is more likely to ensure the evaluation process reflects the goals and objectives of the program.
9. Authentic assessment provides valuable information to teachers on student progress and success of the instruction.

Results were expressed by respondents in line with the characteristics of an authentic assessment that measures all aspects of learning, carried out during and after the learning process, using various tools and resources, the test is only one means of collecting data, tasks correspond to real life, as well as the emphasis on depth knowledge and expertise of learners (Kunandar, 2013).

Obstacles/difficulties encountered in performing authentic assessment is perceived by the respondents. The survey showed that the biggest obstacle is the authentic assessment time. Other barriers perceived by respondents were:

1. The number of students in the classroom too much.
2. Assessment instruments too much and there is no standard.
3. Difficulty making instruments.

4. Administratively add jobs teacher assessment work already a lot.

C. Authentic Assessment in Social Studies Lessons

According to the survey, 100% of respondents express an authentic assessment of learning fits done in social studies. But there are complaints about the materials to be supplied. The material with the interdisciplinary nature makes social studies teacher should be versatile, because the majority of teachers graduated from the courses included in the social sciences instead of social studies education programs. A limited time with a lot of material that makes respondents had difficulty in assessing adapted to the scope and purpose of social studies.

The scope of social studies is a social behavior, economics and culture in the community so that the community be the main source of learning social studies (Winataputera, 2009). The goal is to be able to develop the student as a social actor capable of taking decisions, reasonable, intelligent, committed, responsible and participative. The purpose of learning social studies in accordance with the authentic assessment components such as knowledge, skills, attitudes and values. Based on the scope and purpose of social studies learning is then suitable authentic assessment carried out so as to produce learners that correspond to the learning objectives of social studies. Making the assessment instruments can be simplified and discussed with teachers of other subjects. In practice, teachers tend to think that the implementation of authentic assessment should be like the example given on the results of the training. Though teachers can create their own concept in assessment. Some of the things that teachers can do include:

1. Teachers can utilize the program excel in making the assessment format, making it easy to do the counting each meeting. For teachers who have not mastered the program, given the basic training to be able to perform simple calculations.
2. The format is made as simple as possible, because if it is made of the sheet for one student each meeting will certainly make teachers will be difficult to recapitulate. Examples of formats that can be made:

No	Name	Date..... Theme			Date..... Theme		
		responsibility	confidence	critical	responsibility	confidence	critical

3. At the beginning of the study (recently entered the classroom) teachers can give different assessments for students who are on time and late. Delays are present can be divided again by their old late. The difference in value do not be too far away, for example: "the right time = 100", "late <5 minutes = 95", "Late 5-10 minutes = 90" and so on. Time attendance in class showing discipline, responsible and respect, so that the presence indicator alone can measure multiple competencies attitude.
4. Get used to pray with pronounced each will start the lesson, so that teachers can assess spiritual competence.
5. During the learning process, teachers can make the two formats, one for teachers and one for students. The format for the students can be charged directly by the students concerned. So students also perform the same assessment by teachers against them. Master role to give the knowledge to the students about all the things that exist in the assessment tool so that students can understand how to assess themselves. So the involvement of students in assessment can increase self-confidence.

IV. CONCLUSIONS AND SUGGESTIONS

A. Conclusion

Knowledge of the authentic assessment of teachers is sufficient, but the teacher should be more to learn more about authentic assessment materials and training do the assessment.

Assessment authentic to go through the planning and carried out throughout the learning process and beyond learning. In practical difficulties experienced by teachers will be gradually reduced if teachers are doing the assessments made itself so that its characteristics will be suitable for the material being taught. The advantages of authentic assessment can be a motivation for teachers to use authentic assessment.

Characteristics of social studies learning suited to the characteristics of authentic assessment. Only the necessary simplification and the use of technology to overcome the barriers obtained by the teacher. Student involvement is important in assessing himself, so that teachers can compare the results of the assessment carried out by student teachers.

B. Suggestions

Authentic Assessment is a process that is done continuously and is a unified whole in the process and results. Therefore, a social studies teacher can make this assessment with his own style without departing from the basic principles of authentic assessment.

This study was conducted in respondents is limited, it is advisable to carry out the population and larger samples so the results can be generalized.

REFERENCES

- [1] Arifin, Z. (2009). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosda Karya.
- [2] Azwar, S. (2012). *Sikap Manusia Teori dan Pengukurannya*. Yogyakarta: Pustaka Pelajar.
- [3] Enggarwati, N. S. (2015). Kesulitan Guru SD Negeri Glagah dalam Mengimplementasikan Penilaian Autentik pada Kurikulum 2013. *Skripsi*. Yogyakarta: Fakultas Ilmu Pendidikan Universitas Negeri Yogyakarta.
- [4] Kementerian Pendidikan dan Kebudayaan. (2013). *Ilmu Pengetahuan Sosial: Buku Guru*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- [5] Kunandar. (2013). *Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013)*. Jakarta: RajaGrafindo Persada.
- [6] Kusmijati, N. (2014). Penerapan Penialain Autentik sebagai Upaya Memotivasi Belajar Peserta Didik. *Seminar Nasional Hasil - Hasil Penelitian dan Pengabdian LPPM UMP 2014* (hal. 55-62). Purwokerto: LPPM UMP.
- [7] ampiran 3 Permendiknas No 22 Tahun 2006 tentang Standar Isi. (2006, Mei 23). Standar Kompetensi dan Kompetensi Dasar SMK. Jakarta.
- [8] Lampiran Permendikbud RI No.66. (2013). *Standar Penilaian Pendidikan*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- [9] Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendekia Press.
- [10] Pratiwi, E. S. (2012). Analisis Kesulitan-Kesulitan Guru Dalam Pembelajaran IPS Terpadu (Studi Kasus Pada SMP Negeri 8 Kota Malang). *Skripsi* (hal. <http://karya-ilmiah.um.ac.id/index.php/ekonomi-pembangunan/article/view/20697>). Malang: Fakultas Ekonomi Universitas Malang.
- [11] Widoyoko, S. E. (2007). *Model Evaluasi Program Pembelajaran IPS di SMP*. Diambil kembali dari www.umpwr.ac.id: <http://www.umpwr.ac.id/download/publikasi-ilmiah/Model%20Evaluasi%20Program%20Pembelajaran%20IPS%20di%20SMP.pdf>
- [12] Winataputera, U. S. (2009). *Materi dan Pembelajaran IPS SD*. Jakarta: Universitas Terbuka.
- [13] Zainul, A. (2001). *Alternative Assessment: Applied Approach Mengajar di Perguruan Tinggi*. Jakarta: Ditjen Dikti Depdiknas.

THE IMPLEMENTATION OF ASSESSMENT MODEL BASED ON CHARACTER BUILDING TO IMPROVE DISCIPLINE AND STUDENT'S ACHIEVEMENT

Rusijono

Educational Technology Department, Educational Faculty, Surabaya State University.
rusijonorusijono@gmail.com

Abstract - The purpose of this research were to investigate the effect of implementation the assessment model based on character building to improve discipline and student's achievement. Assessment is part of integral component of learning process so each of instructional process should be done assessment. Assessment model based on character building includes three components, there are: the behavior of student in the instructional process, the effort of student in the instructional process, and student's achievement. This assessment model based on the character building are implemented at epistemology and educational assessment courses, in Graduate Program of Educational Technology Department, Educational Faculty, Surabaya State University. This research used control group pre-test post-test design. The data collecting method was used in this research were observation and test. The observation was used to collect the data about the disciplines of the student in the instructional process, while the test was used to collect the data about student's achievement. Moreover, the study applied t-test to analyze of data. The result of this research showed that assessment model based on character building improved discipline and student's achievement.

Keywords: *Assessment model based on character building, discipline, and student's achievement.*

I. INTRODUCTION

Development of a national character is a fundamental requirement in the process of nation and state, because the only nation that has the strong character and identity will survive at the international level community. Ideologically, the nation 's character building is an attempt to embody the ideology of the nation in the life of the nation. Normatively, national character building is a concrete manifestation efforts to achieve state goals. Generally, the goal of the state are to protect all people and the entire region, promote the general welfare, the intellectual life of the nation, participate in the establishment of world order based on freedom, lasting peace and social justice (Rusijono, 2012)

In the National Education Day anniversary 2010, the President of the Republic of Indonesia launched the implementation of the National Policy for Character Building. The target of this policy are all stakeholders. Specially in National Education Departement, its main focus is on the school (students, teachers, education personnel), family, community (people around the learners), and the environment. Implementation of the National Policy for National Character Building is done gradually and continuously (multiyears). In the future the Ministry of National Education Departement will include character building through the strengthening of the curriculum from kindergarten to university as part of strengthening the national education system. This policy stressed that it will not add subjects matter or new subjects. The implementation of character building will be integrated into existing subjects matter, co-curricular and extracurricular activities, and involves the participation of the environment, families, and communities. (Edi Drajat Wiarto, et al. 2010).

Aristotle stated that good character is a life well-behaved and full of virtue, to behave well towards others (God Almighty, man, and the universe), and against ourselves. Further explained that the characters consist of three inter-related behavioral performance, namely: know the meaning of kindness, willing to do good, and the real well-behaved. According to documents published by Education Ministry of National Education, character building is defined as values education, character education, moral education, which aims to develop the ability of learners to make good decisions, maintain what is good, and realize the goodness at daily behavior (Edi Drajat Wiarto, et al. 2010). Both of the above opinion stated that the character education related to values and behavior. This means that character education is not only related to the knowledge or the cognitive aspects of learners, but also the attitudes and daily behavior of students. As it pertains to daily behavior, so the implementation of character building requires intensive interaction between educators and learners.

The Minister of Education and Culture of Indonesia made a policy on new curriculum, its called Curriculum 2013. One of the principles in this curriculum is developed a balance between the development of spiritual and social attitudes, curiosity, creativity, cooperation with intellectual and psychomotor abilities. The objective of instructional in this curriculum includes attitudes, knowledge, and skills. Qualifications of capabilities associated with behavior that reflects the attitude is the attitude toward religious and social. Qualifications abilities related to the knowledge is factual and conceptual knowledge based on curiosity about science, technology, art, and culture in the knowledge of humanity, national, state, and civilization-related phenomena and events in our homes, schools, and playgrounds. Average qualification capabilities related skills is having the ability to think and act of productive and creative in the realm of the abstract and the concrete in accordance with assigned to him.

Meanwhile, Yahya Umar in Djemari Mardapi (2005) said that efforts to improve education system in Indonesia is not enough focus on input and process of instruction but also focus on controlling the quality of graduates through the development of a standardized assessment system. Djemari Mardapi (2005) also said that assessment has an important role in education, good judgment can encourage students to learn better and encourage educators to teach better. The study, conducted Rusijono, et.al (1999) showed that the patterns of intensity student learning directly related to the assessment system.

Anderson (2003) states the aspects that need to be assessed teachers on learners in the learning process are three things, namely: the behavior of the learning process (classroom behavior), the efforts of students to master competencies (student effort), and the achievements of participants students (student achievement). Behavior in the learning process includes the attitude and actions of students against peers and teachers, particularly the attitude and actions related to the values in the course material and activities in the learning process. Efforts of students refer to the seriousness and discipline of students in instructional process. Discipline and the seriousness associated with affective and psychomotor aspects of the learning process. Learner achievements of learners meant mastery level of competency to be achieved.

These opinion relevance with Koesoema Doni's opinion (2010). He suggested criteria of assessment in the education based on character building, namely: the quantity of attendance, punctuality collect of the task, the number of violence behavior, the participation of student in program of cooperation between schools, student involvement in drugs, academic achievement, and non-academic culture. This opinion tend to the institutional context, because of it, there are several criteria that directly oriented to learner and some to the institution. If these criteria are oriented to learner, the assessment criteria of character education includes the following aspects:

1. Quantity in the presence of learning, including other activities related to learning.
2. Discipline of student in instructional process.
3. The number of violence behavior, drugs, or other improper behavior.
4. Relationships with fellow students, teachers, or others.
5. Academic achievement.
6. Non-academic achievements.

Based on some opinions of the above can be concluded that assessment model based on character building includes three components. First, behavior in the learning process includes the attitudes and behavior against peers and teachers. Attitudes and behavior in this component refers to the values in the material and learning activities. In the context of educational technology, transformation of values can go through the material presented to the students or through learning activities. Second, student effort, it is refer to the seriousness and discipline of

student in the instructional participation. This component includes the following aspects: the presence of learning or lecturing process, the punctuality in instructional participation and collect the assignment, the frequency and quality of question, the frequency and quality of the opinion, the frequency of consultation out of hours lectures in order enrichment, creativity, reasoning that shown in the activities or work-related lectures. Third, the achievement of students, including academic and non-academic.

The implementation of this evaluation model, student must be punctual and active in lectures. The punctuality is assessed in this evaluation model is attendance in the lecture and collect the assignment. In addition, the student must also be active in the learning process because of their activities in the learning process also be assessed. The activity in the learning process comprise the frequency and quality of question, the frequency and quality of the opinion, the frequency of consultation out of hours lectures in order enrichment, creativity, reasoning that shown in the activities or work-related lectures.

Indonesian dictionary (2007), states that discipline with regard to discipline and obedience (compliance) on the rule of order. Meanwhile, Rusdinal and Elizar (2005 : 133) states that discipline can be interpreted as a punishment and training. Discipline is defined as a penalty associated with disorientation behavior that should be relevance with regulation of lectures. Discipline is also defined as training for the process to be a discipline is usually needed for training. In the training, the participants instilled discipline and trained so that they are able to control themselves to act in accordance with the rules set.

Based on the above opinion can be concluded that the discipline is a condition that is created and formed through a process of a series of behaviors that indicate the values of obedience, loyalty, regularity and or orderliness. At the beginning, discipline is a burden for their obligation to act in accordance with the rules. However, when the regulation went right to their soul the discipline of behavior is no longer perceived as a burden, but on the contrary would weigh on him when he did not act in accordance with the rules. The assessment model regulates any behavior of college student, from attendance at lecture, activity in lectures, and the time of the collect assignment. Relevance with the discipline process, at the beginning the evaluation model is certainly burdening but when the regulation in the evaluation model went right to student's soul they are no longer a burden. Since all activities of the students assessed, so it is concluded that the implementation of this evaluation model can improve student discipline.

Domjan (2010: 2) said that "learning is an enduring change in the mechanisms of behavior involving specific stimuli and/or responses that results from prior experience with those or similar stimuli and responses". This definition has a consequence for the study of learning, ie. the evidence of learning is a change in behavior or improving performance. Taylor and MacKenney (2008: 2) have a same opinion, they said that learning is a change in performance through conditions of activity, practice, and experience. This is an operational definition derived, in part, from scientific investigations. In the classroom, the activities and experiences involve telling and listening, judging, reading, reciting, observing demonstrations, experimenting, interacting with pupils and guests, and learning individually. It is concluded that student activity in the learning process will improve learning achievement.

II. RESEARCH METHOD

The purpose of this study is producing assessment model based-on character building to improve discipline and achievement of students. The development model used is ADDIE (Analysis, Design, Development, Implementation, and Evaluation). The analysis phase includes analysis of performance and analysis needs. In this research, the focus of performance analysis is the performance of the students in the lecture. The author's experience shows that the seriousness and discipline students in the lecture still needs to be improved. In addition, the previous academic year's data also shows that learning outcomes were also less. Based on this analysis, it is needed to improve student discipline in the lecture. If the discipline of students increased, it is expected learning outcomes also increased.

The second stage is the design, the step is designed based assessment model of character education that can improve student discipline. Characteristics of the model that is designed to be able to monitor all the activities of students and designing signs so that every student in the lecture activities have contributed to the study results. The third stage is to develop, according to the theory that has been done, the model assessment based-on character building can monitor all the activities of the students in the lecture includes three components. First, the behavior in

the learning process includes the attitudes and behavior against peers and teachers. Second, student effort, it is referred to the seriousness and discipline of a student in the instructional participation. Third, the achievement of students, including academic and non-academic.

The third phase is development. Assessment model which developed based on character building. It is expected improving discipline and achievement of student. Based on the theory that has been analyzed, assessment model developed includes all behavior of student related to the learning process, includes four components, namely: participation in the learning process, assignments, subsumatif and summative examinations. Student participation include: the presence of the lecture, student activity in asking, present their opinion, creativity, reasoning that shown in the activities or work-related lectures, and student activity consult outside lecture hours in order enrichment lecture material. The assignment component include all assignment from the lecture at one semester. The score of this component is mean of all assignment. Subsummative examination administrated two times. Score of the subsummative examination is mean of two times subsummative examination. Then, the final component of this assessment model is summative examination. The weight of participation is two, assignment is three, subsummative is two, and then summative examination is three.

Spector, J. Michael, et.al (2014: 83) said that the duties at the third phase are: develop guidance for the students, develop guidance for the lecturer and conduct formative revisions. The guidance for students and lecture were developed and it was socialized at the beginning of semester. The article of the assessment model based on character building was presented at International Seminar "Sang Guru" September 8th 2012 in Universitas Negeri Surabaya. In this study, the presentation as a formative revisions.

The fourth phase is implementation. The subjects of this study were students of Departement Educational Technology 2014 A and B, Faculty of Education, State University of Surabaya. The number students of Departement were 92 students, it is divided into two classroom, the number of class A 46 students and class B 46 students. The distribution of the classes is not based on the quality of students, so it can be assumed that the quality of students of both classes are equal. So, the implementation of this assesement model used control group pre test post test design.

Subject of study which used in the implementation were Epistemology and assessment. Determination of the subject of study was based on the consideration that the subject of study must be taken by all students of the Faculty of Education. Second, Epistemology was not liked by undergraduate student, while the assessment was liked by undergraduate student. Moreover, the characteristics of them are defferent, the material of epistemology is metacognitive knowledge and the material of assessment is procedural knowledge. Before implementation of assessment model begun, all of the component assessment model were socialized for student.

The variables that observed during implementation of the assessment model were applicability of assessment model, student discipline and student achievement. The indicators of applicability in this study were applicability of participation assessment, assignment, subsummative examination, and summative examination. There was no defference subsummative and summative examination between this assessment model with the previous assessment model, so the component of applicability of the assessment model based on character building which studied just participation component. The indicators of student discipline are attendance punctuality in the lecture and collect the assignment. Then, the indicators of student's achievement were the total score of assignment, subsummative, and summative examination. Participation component is not included in the variable achievement because assessment at control group did not use components participation. Moreover, the study applied t-test to analyze of data.

III. RESEARCH FINDING AND DISCUSSIONS

At the beginning of the lecture, the researcher explains assessment models used on subjects that are being pursued. At that time many questions from students whose goal resist the things that are load on them. This is consistent with the theory that in the early stages of discipline is regarded as a burden, but once explained the advantages of this assessment models eventually they accepted.

The first variable studied in this research is the applicability of the assessment models based on character building. The components of the assessment model are participation in the learning process, assignments, subsumatif and summative examinations. Actually, assignments,

subsumatif and summative examinations had been applied in Graduate Program of Educational Technology Department, Educational Faculty, Surabaya State University. So, there is just one a new component in this assessment model, namely participation. Subcomponent of participation are: the presence of learning or lecturing process, the frequencis and quality of question, the frequencis and quality of the opinion, the frequencis of consultation out of hours lectures in order enrichment, creativity reasoning that shown in the activities or work-related lectures. Complete data on student participation experimental group and the control group can be seen in table 1 below .

Table 1. Participation Data of University Student

Classes	Presence	Question	Opinion	Consult	Creativity	Score of Participation
Experiment	58,5	4,1	3,6	2,3	2,1	71,3
Control	55,7	2,4	1,9	1,7	1,4	64,2

Note: The data is the average of the Epistemology and Educational Assessment

The data in Table 1 indicate that the presence of university student between the experiment class and the control class was almost the same, the average of presence at experiment class is 58.5 and the average at control class 55.7. While the components of the activity of students in asking have a dramatic difference, the average score of the students question in the experiment class was 4.1 and the control class was 2.4. The average score of student express an opinion at the experiment class was 3.6 and the control class was 1.9. The average score of the student activities consulting for enrichment of material at experimental class was 2.3 and the control class was 1.7. While the average score of the creativity of students who appear in asking, questions, or expressing their opinions at experiment class 2.1 and the control class was 1.4. Overall score of student participation at the experiment class was 71.3 and the control class was 64.2.

The results of t test analysis of the participation between the experiment class and a control class was 4.78. This result is greater than the $t_{table} = 2,021$ (df 44 and the significance level 0.05). Based on the results of this analysis we can conclude two things. First, a model-based assessment of character education can be implemented properly. Implementation of this assessment models require lecturers to know students well because he had to record every student who ask or express the opinions. If the number of students in one class are too large it is difficult to be able to carry out this assessment model. Secondly, there is a significant difference between the participation of students in the experiment class and the control class.

The second variable which studied in this research was the discipline of students in the lecture process. There were two components of discipline, namely: attendance punctuality in the lecture and collect the assignment. The data of the two components can be seen in table 2.

Table 2. The Discipline Data of University Student

Classes	Puncuality		Skor Rata-rata Kedisiplinan
	Attendance in the Lecture	Collect the Assignment	
Experiment	97,5	93.7	95.6
Control	92,8	76.8	84.8

Note: The data is the average of the Epistemology and Educational Assessment

The results of t test analysis of the discipline students between the experimental class and a control class was 5.21. This result is greater than the $t_{table} = 2,021$ (df 44 and the 0.05 significance level). Based on the results of this analysis can be concluded that there was significant differences between class that implement the assessment model based on character building with it not implement the assessment model based on character building. Variable discipline is a variable that is embedded in the lecture so it is not possible to do a pretest. Therefore, as a substitute for pretests used data on experimental class student discipline in the course of research methodology. The average score of student discipline in following the course of research methodology was 84.4. This score is slightly below the average score of classroom discipline control.

The results of t test analysis of the experiment class score student discipline in following the course of research methodologies and the two subjects tested (Philosophy of Science and Evaluation) was 5.61. This result is greater than the $t_{table} 2,021$ (df 44 and the 0.05 significance level). Based on the results of the above analysis it can be concluded that the implementation of assessment model based on character building can improve student discipline in following the lecture.

The third variable was studied in this research is student achievement. The component of achievements include score assignments, score of subsummative and summative examinations. The data of the three components can be seen in Table 3 .

Table 3. The Achievement Data of University Student

Classes	The Component of Achievement			Score of Achievement
	Assignment	subsummative	summative	
Experiment	78	81	80	79,7
Control	76	74	73	74,2

Note: The data is the average of the Epistemology and Educational Assessment

The results of t test analysis of student achievement between experimental class and control class 3.11. This result is greater than the ttable 2,021 (df 44 and the 0.05 significance level). This study does not do the pretest, as a substitute for pretests was used student achievement index in the previous semester. The results of t-test between achievement student on epistemology and assessment in a control class with student achievement index in the previous semester was 1.99 (< ttable = 2.021). Based on the results of the two analysis it can be concluded that the implementation of assessment model based on character building can improve student achievement.

IV. CONCLUSION

Based on the analysis that has been described, it can be drawn the three conclusion. First, a model-based assessment of character education can be implemented properly. Implementation of this assessment models require lecturers to know students well because he had to record every student who ask or express the opinions. Therefore, this assessment model only suitable for class sizes relatively small. Second, the implementation of assessment model based on character building can improve student discipline in following the lecture. Third, the implementation of assessment model based on character building can improve student achievement

REFERENCES

- [1] Abdullan Munir (2010). Pendidikan karakter, membangun karakter anak sejak dari rumah. Yogyakarta: Pusataka Insan Madani.
- [2] Anderson, Lorin W. (2003). Classroom assessment, enhancing the quality of teacher decision making. Marwah: Lawrence Erlbaum Associates, Publishers.
- [3] Anderson, O.W. dan Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing. New York: Addison Wesley Longman, Inc.
- [4] Departemen Pendidikan Nasional Republik Indonesia (2005). Peraturan Pemerintah Republic Indonesia Nomor 19 tahun 2005 tentang Standar Nasional Pendidikan. Jakarta: Badan Standar Nasional Pendidikan.
- [5] Djemari Mardapi. (1999). Azas performance-based evaluation. Makalah disampaikan pada Lokakarya "Performance-Based Evaluation dan Bank Soal" tanggal 18 – 19 Nopember 1999.
- [6] Djemari Mardapi. (2001). Evaluasi proses & hasil belajar: Asesmen alternatif. Makalah disampaikan pada tanggal 16 Agustus 2001.
- [7] Doni Koesoema A. (2010). Pendidikan karakter, strategi mendidik anak di zaman global. Jakarta: Gramedia.
- [8] Edi Drajat Wiarto, dkk. (2010). Pembangunan karakter bangsa tahun 2010 – 2025. Jakarta: Pemerintah Republik Indonesia.
- [9] Gronlund, N. E. (1985). Measurement and evaluation in teaching (6th ed.). New York: Macmillan Publishing Company.
- [10] Johnson D.W. dan Johnson R.T. (2002). Meaningful assessment. Boston: Allyn and Bacon.
- [11] Reynolds, C.R., Livingstone, R.B., & Willson, V. (2010). Measurement and assessment in education. Second Edition. London: Pearson Education International.
- [12] Rusdinal dan Elizar (2005). Pengelolaan kelas di Taman Kanak-kanak. Jakarta: Direktorat Pembinaan Pendidikan Tenaga kependidikan dan Ketenagaan Perguruan Tinggi.
- [13] Rusijono, dkk. (1999). Pengembangan system evaluasi hasil belajar untuk meningkatkan prestasi belajar mahasiswa IKIP Surabaya. Laporan Penelitian. Surabaya: Lembaga Penelitian IKIP Surabaya.
- [14] Rusijono (2009). Pelaksanaan system penilaian perkuliahan Universitas Negeri Surabaya di Fakultas Ilmu Pendidikan. Laporan penelitian. Surabaya: Lembaga Penelitian Universitas Negeri Surabaya.
- [15] Rusijono (2011). Implementation of character building in Distance Learning at Universitas Terbuka, Article presented at The 24th Conference International Council for Open and Distance Education in Bali 2 – 5 October 2011.
- [16] Rusijono (2012). The Development Assessment Model Based on Character Building. Article Presented at International Seminar on September 8th 2012 at Universitas Negeri Surabaya.
- [17] Stufflebeam, D.L. dan Shinkfield, A.J. (1985). Systematic evaluation. Boston: Kluwer-Nijhoff Publishing.

-
- [18] Taylor, G.R. & MacKenney, Loretta (2008). Improving Human Learning in the Classroom, Theories and Teaching Practices. Lanham: Rowman & Littlefield Education.
- [19] Tilstone, C. (2002). Observing teaching and learning. London: David Fulton Publishers.

THE DESIGN OF PERFORMANCE ASSESSMENT BASED GUIDED INQUIRY FOR EMPOWERING STUDENTS' ARGUMENTATION SKILLS

Saiful Imam Ali Nurdin¹, Viyanti²

¹Teacher Senior High School YP Unila

¹saiful.imam06@gmail.com

Abstract - This research is a preliminary stage in developing performance assessment based (PABs) guided inquiry to empower students argumentation skills. The purpose of this research are: (1) determine the needs of teachers and students will be Performance Assessment. (2) designing Performance Assessment can be used to empower the student argumentation skills. The preliminary study was conducted in public and private high school in the city of Bandar Lampung in the first semester of the academic year 2015/2016 using random sampling technique. The research subject of high school students at different levels numbered 80 students and 6 physics teachers. The data were obtained using a questionnaire and literature review. The analysis carried out by stages: coding, tabulation of data, data analysis, design of Performance Assessment obtained from the literature review, interpret the results of the analysis in accordance with the problems and research questions and make their conclusions. The findings obtained in this preliminary study are: 1) Teachers and students in public and private high school in the city of Bandar Lampung requires an assessment instrument in the form of performance assessment. 2) Design of Performance Assessment can be used to empower the student argumentation skills. Design Performance Assessment consists of several stages consisting of orientation, exploration, formation of concepts, applications, cover and display argumentation skills. At every stage is an assessment technique and different instruments.

Keywords: PABs, guided inquiry, argumentation skills.

I. INTRODUCTION

Assessment is one of the important components in the learning process. Assessment is an overview of the evaluation of the students in understanding the learning process that has been given by the teacher in the school also as an answer of the question of how well the learning outcomes of students in the learning process. Assessment also provides insight and feedback for learners to find achievement indicators tested by the teachers. Results of the assessment also can be used as a stuff to improve the learning process that has been implemented. Assessment can describe the development of learners, data also can be used as an evaluation.

[1] Standard Assessment Education is the criterion on the mechanisms, procedures, and instruments of assessment of the learning outcomes of learners. Rate of education as a process of collecting and processing information to measure the achievement of learning outcomes of students include: authentic assessment, self-assessment, portfolio-based assessment. The assessment process is often used by teachers is still limited to daily tests, midterm replay and replay the end of the semester that overall only limited emphasis on cognitive aspects and doesn't take the advantage of other types of assessments which have been stipulated in the existing legislation including authentic assessment. Authentic score is one of type assessment that is recommended in the learning process that can improve the student learning outcomes from both aspects of cognitive affective, and psychomotor. Authentic assessment covers several types of assessments that is performance assessment, project assessment, portfolio assessment, self-assessment, assessment of peers, and written assessment [2].

The type of authentic assessment that will be developed as an instrument in development research writer is an instrument of performance appraisal. The performance assessment is

defined as concrete and authentic tasks that require students to do something with the knowledge and skills, such as giving an activity, demonstration, presentation or write a report [3]. Performance assessment can help students to develop the potential of such investigation skills, problem solving, verbal presentations, organizing and writing skills [4]. Performance assessment involves students to demonstrate their ability to think, to do a specific skill or to create a particular product [5]. Assessment of performance need to be developed in the learning process, being able to explore the potential and talent possessed by learners. [6] is obviously a performance based assessment approach is needed to assess the gifted students in learning and performance tasks to challenge gifted learners to express intellectual capacity. Performance Based Assessment (PBAs) has been successfully used to measure complex reasoning, think critically, and content learning in science. Therefore, when designing a complex and rigorous curriculum for talented students, who have some learning outcomes associated with higher thinking level, the contents of the progress of science, and a conceptual understanding, some performance measures must be embedded [7]. The performance assessment is also used to cultivate the talent possessed by the students can be used also in measuring high-level thinking skills. [8] assessment used to evaluate the performance of higher-level thinking and the acquisition of knowledge, concepts and skills needed for students to succeed in the workplace of the 21st century. The performance assessment should be carried out to the maximum by a teacher to make all latent potential on students exist also be seen observed and can be evaluated objectively.

Implementation of learning process can be use some learning models for the achievement of learning goals. From some of existing models one of which is a model of inquiry learning. Inquiry learning model emphasizes the activity of a student in the search of information either by observation or experiment. The application of the model of inquiry is important in the learning process. Model of inquiry effective to assess student performance assessment.[9] the application of guided inquiry learning model effectively to improve the mastery of concepts and performance. Learning by using inquiry learning model will bring multiple skills which present a phenomena, making research questions, write the hypotheses, planning experiments to test the hypothesis, after obtaining the data, analyze and interpret the results, draw the conclusions, and also express the opinions [10]. The inquiry process in learning are produce skills that can be measured using the Performance Assessment.

Learning to be expected from the implementation of the curriculum in 2013 include the involvement of the student in the learning process. Teachers only act as a facilitator and provide direction and motivation. The ability of the student is expected to appear a lot of activity in the learning ability to deliver such an argument when interact in the learning activities. [11] there are several strategies to build the student argument are: to explain what evidence to support each theory, construct an argument using a pattern structured that included questions guiding, predicts the experimental results, based on the right arguments, observing experiments and explains the results (predict, observe, explain), to design experiments, execute and discuss the results. Another thing that can be done to foster among students, according argument [12] infuse meaning and essence of the scientific content developed a scientific concept would be a way to formulate an argument. Learning environment led to participate in argumentation students, according to Jimenez (2008) in [13], the characteristics of an optimal learning environment to develop arguments relating to students, teachers, curriculum, assessment, reflection, and communication is as follows: (1) The student must be active in the learning process; they should assess the knowledge, build their claims, and was critical of other people; (2) the teacher should adopt to a student-centered learning, acting as a role model of how they verify their claims, supports the development of understanding of the nature of knowledge among students, and to adopt learning strategies such investigation; (3) The curriculum must include an authentic problem-solving approach, which would require students to learn by inquiry; (4) students and teachers should be skilled in assessing claims, and assess students must go beyond a written test; (5) The student must be reflective about their knowledge and understanding of how it was obtained, and finally (6) the student should have the opportunity to have a dialogue in which cooperative learning will take place.

Based on the description above, the purpose of the research that will be held by the researchers include: (1) Determine the needs of teachers and students will be Performance Assessment. (2) Designing Performance Assessment can be used to empower the student argumentation skills.

II. RESEARCH METHODS

This study is a preliminary study. The sampling technique that will be use in this study is random sampling that is considered of interest to obtain data on the learning process. Subjects are 6 physics teacher and 80 students in Senior High School and Private School in Bandar Lampung in academic year 2015/2016. The data collection is using a questionnaire analysis of teachers and students. The data is processed and analyzed with the following steps: The first step is to code the data from the results of a questionnaire distributed, the second step is data tabulation for ease in reading, categoring, and analyzing; The third step is the qualitative data analysis is to analyze the data by breaking and linking of data and information relating to the focus of the research, and the next step is to interpret the results of the analysis in accordance with the problems and research questions and make the conclusions.

III. RESULTS AND DISCUSSION

The results of the analysis of preliminary research that has been done show in several table's below.

Table 1. Implementation of Performance Assessment According to Teacher

No	Analysis Questions
1	67% of teachers use authentic assessment tools to assess student learning outcomes
2	83.33% does not make any assessment in accordance with the curriculum in 2013 to assess student performance
3	50% of teachers has make the performance assessment device refers to the specific learning model in ratings
4	50% of teachers never use the device performance assessment refers to the specific learning model in ratings
5	50% of teachers have difficulty in making the device performance assessment refers to the specific learning model

Based on the analysis in Table 1, there are some teachers do not use performance assessment instruments and experiencing difficulties on its development. Implementation of performance assessment according to students in Table 2.

Table 2. Implementation of Performance Assessment According to Students

No	Analysis Questions
1	73% of physics teachers always make an assessment in the classroom when the learning process for the aspects of attitudes, knowledge, and skill shown by yourself and your friends
2	18% assessment conducted by the Physics teacher only in the form of a written test
3	68% of the value given by a teacher in physics has been fair with capabilities
4	64% of students received grades in physics based on the overall aspects of the set value
5	69% physic teachers tells about the aspects to be assessed and how the assessment to all students at the beginning of learning process
6	5% of students saw the Physics teacher holding the assessment sheet while walk around in the class room during the learning process
7	64% Physics teacher asks students to demonstrate performance or skill in teaching later assessed by the teacher
8	65% of students always showed a good performance during the learning to get the best score in physics

Based on the analysis in Table 2 there is something contrary between the answers of number 6 question with answer number 8. In number 8 answer, students said they always show the performance in learning, but for number 6 answer, students who saw the teacher bring the score instrument. Every learning are little, but in the implementation, the teacher performance cannot be separately from scoring sheet in their learning process. This may happen due to lack understand of students on performance score itself. While, analysis of performance needs assessment according to teacher in Table 3.

Table 3. Teachers' Performance Assessment Needs

No	Analysis Questions
1	100% of teachers willing to develop the performance assessment based guided inquiry to empower students argumentation skills
2	100% of teachers agree that developed the device performance assessment based guided inquiry to empower students argumentation skills

Based on the analysis in Table 3, it can be concluded that it takes develop performance assessment based guided inquiry to empower students argumentation skills. While, analysis of performance needs assessment by students presented in Table 4.

Table 4. Students Performance Assessment Needs

No	Analysis Questions
1	64% of students agreed that teachers develop their physical performance assessment tools
2	68% of students agreed that teachers of physics apply to the performance assessment tools in the classroom

Based on the analysis in Table 4, it can be concluded that the majority of students agreed in the learning process teachers use performance assessment.

Assessment in the learning process using performance assessment is very important because it can assess talented students and capable of measuring the ability in authentic of the student itself. This is in line with the result of research [14] assessment is uses to assess achievement the competence that requires student to specific task like: laboratory, experiments, presentations, discussions, role playing, and others. How this assessment is considered more authentic than the written test for what was rated better reflect the actual ability of learners. The performance assessment can also be used to assess psychomotor abilities of students in learning activities including lab. Practicum itself important in the learning process to make the students better understand the material given by the teacher. This is in line with the opinion [15] the use of rubrics Performance Assessment can also help in the assessment of performance (psychomotor) student can practice skills and students' performance in the lab.

The use of models in learning inquiry is expected to bring up the argumentation skills of students and students can gain the maximum of learning results. [16] due to the improvement of student learning outcomes assessment characterizes authentic inquiry model was developed to train students actively in the learning process is challenged to find the problem and solved by finding a solution to a problem that is faced in the real world kontesks. Model of inquiry is an appropriate alternative because it can increase the performance of students with significant according to research conducted [17] which states students by using guided inquiry showed better performance. Inquiry model will be able to help students achieve the learning objectives of science among students will build the concept, the ability to think, science process skills, argumentation, habits of mind, and understanding the nature of science [18]. In the barrel with Nuangchalerm according to research [19] which states that learning by inquiry will make students more focused and stronger in the achievement of the argument.

Model inquiry used by researchers to use the steps proposed [20] as follows:

A. Orientation

In the orientation phase is done is raise the interest of students to the learning process, provide motivation, generating curiosity, and build new information with prior knowledge.

B. Exploration

In the exploration phase provides opportunities for students to make observations, collect and analyze information, and develop hypotheses based on the problems posed teacher.

C. Concept Formation

In this phase, a follow up of the exploration stage which requires students to find the relationships between concepts and encourage students to think critically and analytically to build a conclusion.

D. Application

The concept of new knowledge that has been obtained in apply in situations such as exercise

E. Closure

Phase cover (closures) directs the students to be able to report the result, and reflecting on what they have learned, to consolidate knowledge.

Design performance assessment which will be developed in addition to using models termbimbing inquiry in the learning process, the researchers are also trying to grow the skills of argumentation from the students is one of the skills contained in the learning objective science. The use of components that exist on desain arguments later refers to the opinion Toulmin 1958 in [21] which states that in building the skills of the student's argument has there are three basic components: a claim, evidence, and warrants. The explanation for each of three components of which the claim is a statement of a person who shows it to believe it was true, for example, conclusions, answers to questions or problems. Evidence is scientific data to support the claim. Scientific data consisting of information, such as observation and measurement. Warrants or a denial that makes certain claims incorrectly by to use additional evidence and reasoning to justify it.

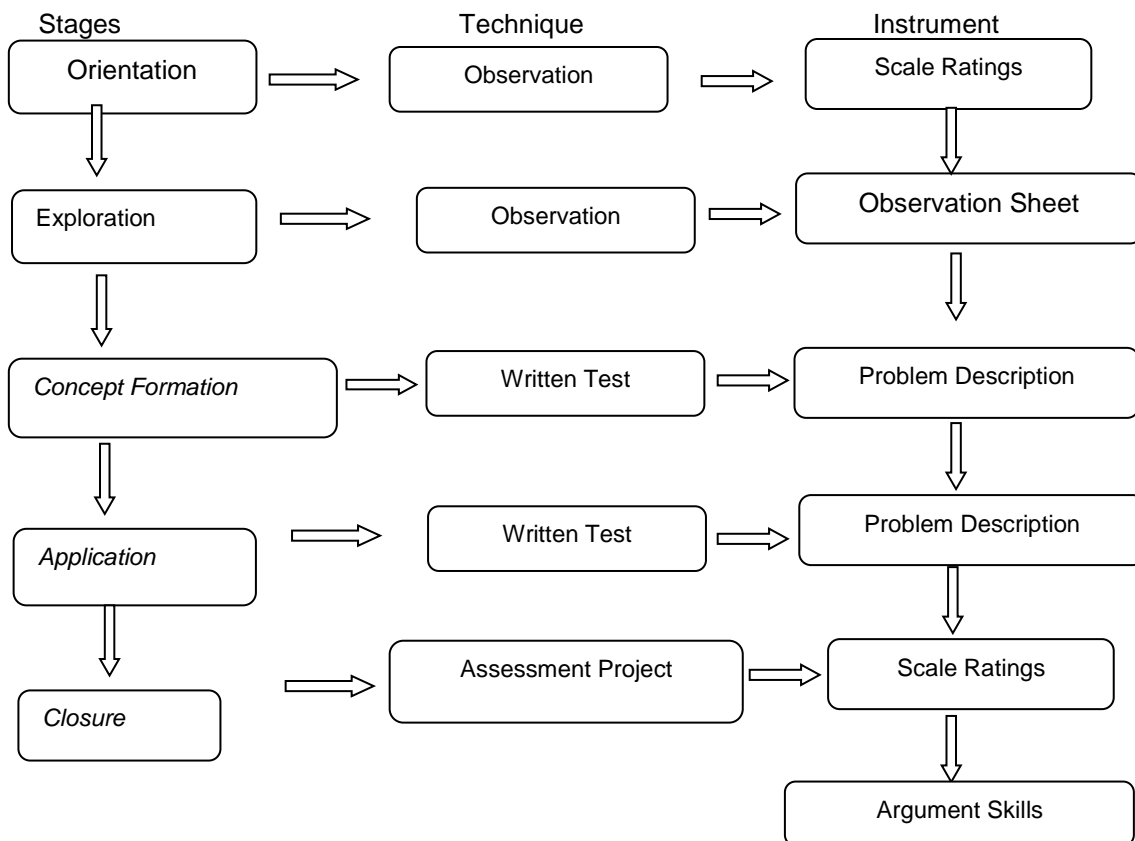


Figure 1. Design of Performance Assessment based guided inquiry argumentation skills to empower students

IV. CONCLUSION

Based on studies in this preliminary study can be summarized as follows: (1) teachers and students in public and private high school in the city of Bandar Lampung requires an assessment instrument in the form of performance assessment, (2) In most design Performance Assessment can use to empower the skills of the argument of students. Suggested guided inquiry performance assessment charged to develop the skills of argument that will be designed to systematics as follows: orientation, exploration, formation of concepts, applications, cover and display argumentation skills.

ACKNOWLEDGMENT

This project would not be possible without the help from members of our team who are not author on this paper Asih Sulistia Ningrum, Erlida Amnie, Luthfi Riadina, and Feryco Candra.

REFERENCES

- [1] Permendikbud No 69. 2013. Standar Penilaian Pendidikan. Jakarta: Depdiknas.
- [2] Kunandar, Authentic Assessment, Jakarta: Rajawali Pers, 2013.
- [3] A. Nitko, Educational assessment of students, Englewood Cliffs: Prentice-Hall, 2004.
- [4] M. Metin, "Performans değerlendirmenin öğretmen adayları üzerindeki etkisinin incelenmesi," Çağdaş Eğitim Dergisi. Turki, pp 28-35, 2008.
- [5] M. Metin, and H. Ozmen, "Investigation of Teacher Opinions about Performance Assessment with Respect to the Gender and Branch Variables," Turkish Science Education. Turki, Vol 4, 2011.
- [6] J. VanTassel-Baska, "Performance Based Assessment The Road to Authentic Learning for the Gifted," Journals permissions. USA, Vol 37, 2013.
- [7] K. Kim, J. VanTassel-Baska, B. Bracken, A. Feing, and T. Stambaugh, "Assessing Science Reasoning and Conceptual Understanding in the Primary Grades Using Standardized and Performance Based Assessments," Journal of Advanced Academics. USA, Vol.25, pp 47-46, 2014.
- [8] D. Wren, "Performance Assessment: A Key Component Of A Balanced Assessment System," Research Brief. Report From The Department Of Reseach Evaluation, and Assessment. USA, pp 2, 2009.
- [9] L. Praptiwi, sarwi, and L. handayani, "Effectiveness of Guided Inquiry Learning Model Experiments Assisted My Own Dictionary To Improve Performance Mastery Concept And RSBI junior high school students," Unnes Science Educational Journal.Indonesia, Vol. 2, pp. 87-95, 2012.
- [10] D. Katchevich, A. Hofstein, and R. Naaman, "Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments," Research Sains Education. Israel, pp 317-345, 2011.
- [11] D. Katchevich, "The Characteristics of Open – Ended Inquiry-Type Chemistry Experiments That Enable Argumentative Discourse," Sisyphus Journal of Education. Israel, Vol. 2, pp 74-99, 2014.
- [12] S. Erduran, S. Simon, and J. Osborne, "Tapping into argumentation: developments in the application of Toulmin's Argument Pattern for studying science discourse," Science Education. Vol 88, pp 915-933, 2004.
- [13] D. Katchevich, "The Characteristics of Open – Ended Inquiry-Type Chemistry Experiments That Enable Argumentative Discourse," Sisyphus Journal of Education. Israel, pp 74-79, 2014.
- [14] A. Samsudin, E. Suhendi, R. Efendi, and A. Suhandi, "Development of "CELLS" in Basic Physics Experiment for Developing Performance Skills and Improve Student Learning Motivation," Jurnal Pendidikan Fisika Indonesia. Indonesia. Indonesia, Vol. 8, pp. 15-25, 2012.
- [15] N. Puspitasari, S. S. Haryani, and N. N. Widiarti, "Development of Performance Assessment Rubric At Salt Hydrolysis Practicum," Jurnal Inovasi Pendidikan Kimia. Indonesia, Vol. 8, pp. 1250-1259, 2014.
- [16] S. Triamijaya, and S. haryani, "Development on Authentic Assessment by Inquiry-based Material Objects Classification," Unnes Science Education Journal.Indonesia, Vol. 4, 2015.
- [17] I. Bilgin, I. "The effects of guided inquiry instruction incorporating a cooperative learning approach on university students' achievement of acid and bases concepts and attitude toward guided inquiry instruction," Scientific Research and Essay. Turki, Vol 4, pp 1038-1046.2009.
- [18] P. Nuangchalem, "Inquiry based Learning in China Lesson learned for School Science Practices," Asian Social Science. Thailand, Vol. 10, 2014.
- [19] D. Katchevich, A. Hofstein, and R. Naaman, "Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments," Research Sains Education. Israel, pp 317-345, 2011.
- [20] Hanson, D. M. 2012. Designing Process-Oriented Guided-Inquiry Activities. (access from [http://quarknet.fnal.gov/fellows/TLDownloads/Designing POGIL Activities.pdf](http://quarknet.fnal.gov/fellows/TLDownloads/Designing%20POGIL%20Activities.pdf)).
- [21] D. Katchevich, A. Hofstein, and R. Naaman, "Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments," Research Sains Education. Israel, pp 317-345, 2011.

THE INFLUENCE OF CLASS CLIMATE AND SELF CONCEPT TOWARDS ACHIEVEMENT MOTIVATION AND PHYSICS LEARNING RESULT OF STUDENT AT XI IPA GRADE SMA NEGERI 1 KAHU

Satriani ¹⁾, Kaharuddin Arafah ²⁾, Muris ³⁾

¹ Student at Phisycs Education Post Graduate Program, State University of Makassar

² Lecturer at Phisycs Education Post Graduate Program, State University of Makassar
satry_icp09@yahoo.com

Abstract - This research is expost-facto research which aims to determine the influence of: (1) the class climate towards physics learning result, (ii) class climate towards achievement motivation, (iii) self concept towards physics learning result, (iv) self concept towards achievement motivation, and (v) achievement motivation towards learning result. The population in this study were all students of XI IPA Grade SMA Negeri 1 Kahu as many 181 students. The sample taken by using Slovin technique with 126 students. The process of collecting data using questionnaires and test of physics learning result that have been tested empirically. Data of the research result were analyzed by using analysis method of Structural Equation Modeling (SEM) with technique of Analysis of Moment Structures (AMOS). The analysis procedure is performed with descriptive analysis and inferential analysis, factor analysis and verification of structural model AMOS. The result of research showed that the structural equation model that describe the influence of class climate, self concept, achievement motivation, and physics learning result can be accepted. Through the model can be conclusion that: (i) class climate doesn't have positive direct influence and significant towards physics learning result, (ii) class climate has positive direct and significant toward achievement motivation, (iii) self concept doesn't have positive direct influence and significant towards physics learning result, (iv) self concept has positive direct and significant toward achievement motivation, (v) achievement motivation doesn't have positive direct influence and significant towards physics learning result

Keywords: *Expost-Facto, Class Climate, Self Concept, Achievement Motivation, Physics Learning Result*

I. INTRODUCTION

In constitution about national education system "National education serves to develop the ability to form the character and civilization of the nation's dignity in the context of the intellectual life of the nation, is aimed at developing students' potentials in order to become a man of faith and fear of God Almighty, the noble character, healthy, knowledgeable, skilled, creative, independent and become citizens of a democratic and responsible ". Potential students as a subject of study as well as the role of the teacher in the learning process is a very important key in determining the success of teaching in schools. The success of teaching can be seen from the results of their study. The results of this study are determined by internal and external factors such as the learning environment, classroom and school climate, academic self-concept and motivation of learners.

Judging from the role of teacher as an educator, mentor, coach, and leaders who can create a classroom climate that is attractive, safe, and comfortable, its presence in the midst of students can break the ice rigidity, stiffness and boredom learning droop received by the students. Class climate is not conducive to negative affect the learning process and the difficulty of learning objectives achieved. Classroom climate and attract much needed conducive to learning physics

largely considered difficult by students so that the taste is hard it can be minimized with the comfort in the classroom. Students who have a positive perception of the classroom climate will feel comfortable when entering the classroom, knowing that there will be the care and respect them, and believe that it will learn something valuable. But otherwise the students who have the perception of the classroom climate that is negative student will feel fear if they are in the classroom and doubt whether they will get valuable experience. This is similar to the research that has been done by (Hadinata, 2009) which found that classroom climate contribute significantly to the motivation to learn in high school students [1].

In addition to classroom climate there are also internal factors that influence learning outcomes ie self-concept. According to (Desmita 2010) mentions that self-concept is the notion of self that includes beliefs, views, and votes over oneself. Students who are anxious in the face of the final exam by saying "I actually stupid boy, I certainly would not get a good value", is already reflecting expectations of what will happen with the test results. The phrase indicates his belief that he does not have the ability to obtain a good value [2].

All of these things must be accompanied also with motivation. Motivation is very important in teaching belajar activities, because their motivation to encourage the spirit of learning and conversely lack of motivation will weaken the spirit of learning. A student who learned without motivation or lack of motivation, will not work with the maximum. Thought is supported by studies that have been done before. As research has been done by (Saadi, 2012) that the achievement motivation have a relationship that is positively correlated to the learning result.

Thus the rationale encourage researchers to conduct research on the influence of class climate and self-concept towards achievement motivation and physics learning result of students at XI IPA Grade SMA Negeri 1 Kahu. The problems in this research are: (1) Does the class climate have positive direct influence towards physics learning result? (2) Does class climate have positive direct influence towards achievement motivation? (3) Does the self-concept have positive direct influence towards physics learning result? (4) Does self-concept have positive direct influence towards achievement motivation? (5) Does the achievement motivation have positive direct influence towards physics learning result? The purpose of this study was to answer the formulation of the problem, namely (1) to determine the direct influence of class climate towards physics learning result, (2) to determine the direct influence of class climate towards achievement motivation, (3) to determine the direct influence self-concept towards physics learning result, (4) to determine the direct influence of self-concept towards achievement motivation, and (5) to determine the direct influence of achievement motivation towards physic learning result. The results of this study are expected to provide good benefits for reference to learners to be able to improve learning outcomes through the establishment of positive classroom climate, form a good self-concept along with achievement motivation.

II. RESEARCH METHOD

The type of research is the study "ex post facto", which is causality and correlation. This study attempted to investigate the direct influence of the independent variables namely the class climate and self-concept on physics learning result as dependent variables, both directly and through the achievement motivation as an intervening variable. Designs of linkages between these variables are described as follows.

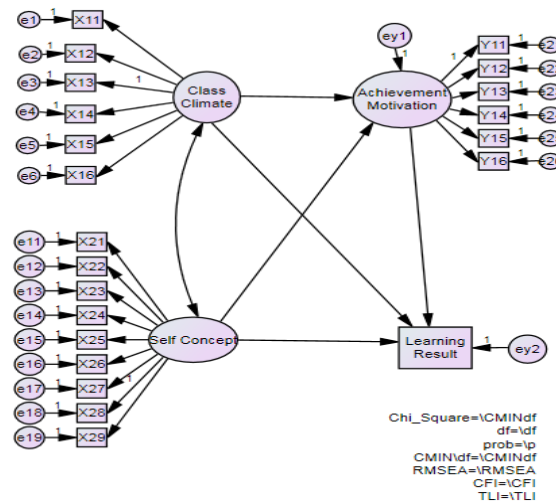


Figure 1. Structural Model of Functional Interwoven Proposed Among Variables Research

The populations in this study were all students of class XI IPA SMA Negeri 1 Kahu academic year 2015/2016, Bone Regency, and South Sulawesi. Distribution of students in each class is shown in table 1 below.

Table 1. Distribution Of Student Class XI IPA

Name of Class	Number of Students
XI IPA 1	31
XI IPA 2	29
XI IPA 3	29
XI IPA 4	30
XI IPA 5	32
XI IPA 6	30
Total	181

The technique used to determine the sample size is by slovin technique. As for the distribution of the sample is more presented in table 2 below.

Table 2. Distribution Of Research Sample

Name of Class	Number of Students
XI IPA 1	20
XI IPA 2	-
XI IPA 3	29
XI IPA 4	25
XI IPA 5	25
XI IPA 6	27
Total	126

The instrument used in this study consisted of questionnaire of the class climate, self concept questionnaire, questionnaire achievement motivation, and test of physics learning result.

Before the instrument is ready for use, it must first be validated instrument that is validation of content and empirical validity. Contents validation test conducted on measuring instruments. Analysis of the contents of the instrument validation is done in this research that uses models Gregory in the form of a model agreement among experts. Empirical validation test performed on the test results the instrument consisting of item validity test and reliability test. Types and data collection techniques used in this study can be seen in the following table.

Table 3. Types and Data Collection Techniques

Data	Type of Data	Data Collection Techniques	Data Sources	Instrument
Class Climate	Interval	Questionnaire	Students	Questionnaire of class climate
Self Concept	Interval	Questionnaire	Students	Questionnaire of self-concept
Achievement Motivation	Interval	Questionnaire	Students	Questionnaire of Achievement Motivation
Physics Learning	Interval	Questionnaire	Students	Questionnaire of Physics

Result				Learning Result
--------	--	--	--	-----------------

Analysis of the data used in this study consisted of a statistical analysis of descriptive and inferential analysis to test the hypothesis.

a. Analysis of Descriptive Statistic

A function of descriptive analysis is to provide an overview of the data obtained, such as: the number, maximum, minimum, mean, mode, median, standard deviation and variance.

b. Normality Test

Normality test is used to determine whether the data sample studied came from populations with normal distribution or not. Values were considered in determining test multivariate normality is critical ratio by using AMOS 22.0.

c. Linearity Test

Linearity test is performed to determine whether there is a linear relationship between the independent variables with the dependent variable using SPSS 22.0. The variable is said to be linear with other variables when $F_{\text{calculate}} < F_{\text{table}}$

d. Multicollinearity Test

Multicollinearity test is performed to determine whether there is a significant relationship between independent variables in a multiple linear regression model. Multicollinearity test was performed using statistical program SPSS 22.0 to see the value of Tolerance and Variance Inflation Factor (VIF) in the regression model.

e. Factor Analysis

Factor analysis was performed using AMOS 22.0 to be able to test the influence of the indicator with latent variables, a model must be eligible Goodness of Fit, which is an index that is used as a reference model is said to be acceptable fit. The index used is the Chi-square, CMIN / df, TLI, CFI and RMSEA.

III. RESULT AND DISCUSSION

A. Analysis Result of Research Data

1. Descriptive Statistic

Table 4. Summary of Descriptive Analysis Result

Statistic	Achievement motivation	Class Climate	Self Concept	Physics Learning Result
Number of responden	126	126	126	126
Mean	143.8095	135.5952	136.9921	14.2302
Std. Error of Mean	1.00764	.97721	1.17419	.30093
Median	145.0000	135.0000	138.0000	15.0000
Mode	153.00	130.00 ^a	140.00	15.00
Std. Deviation	11.31068	10.96918	13.18029	3.37796
Variance	127.931	120.323	173.720	11.411
Skewness	-.271	-.114	-.373	-.253
Std. Error of Skewness	.216	.216	.216	.216
Range	62.00	53.00	75.00	18.00
Minimum	109.00	109.00	96.00	3.00
Maximum	171.00	162.00	171.00	21.00
Sum	18120.00	17085.00	17261.00	1793.00

Research result data of variable the class climate then presented the list of frequency distribution as shown in table 5 below.

Table 5. Distribution Of Frequency, Percentage, And Categories For Class Climate

No	Interval Score	Frequency	Percentage (%)	Category
1	36 – 64	0	0	Very Low
2	65 – 93	0	0	Low
3	94 – 122	15	11.91	Moderate
4	123 – 151	102	80.95	High
5	152 – 180	9	7.14	Very High
Sum		126	100	

Research result data of variable the self concept then presented the list of frequency distribution as shown in table 6 below.

Table 6. Distribution Of Frequency, Percentage, and Categories for Self-Concept

No	Interval Score	Frequency	Percentage (%)	Category
1	38 – 68	0	0	Very Low
2	69 – 99	1	0.79	Low
3	100 – 130	32	25.40	Moderate
4	131 – 161	90	71.43	High
5	162 – 192	3	2.38	Very High
Sum		126	100	

Research result data of variable the achievement motivation then presented the list of frequency distribution as shown in table 7 below.

Table 7. Distribution Of Frequency, Percentage, and Categories for Achievement Motivation

No	Interval Score	Frequency	Percentage (%)	Category
1	36 – 64	0	0	Very Low
2	65 – 93	0	0	Low
3	94 – 122	4	3.18	Moderate
4	123 – 151	87	69.04	High
5	152 – 180	35	27.78	Very High
Sum		126	100	

Research result data of variable physics learning result then presented the list of frequency distribution as shown in table 8 below.

Table 8. Distribution Of Frequency, Percentage, and Categories for Physics Learning Result

No	Interval Score	Frequency	Percentage (%)	Category
1	0 – 5	1	0.79	Very Low
2	6 – 11	25	19.84	Low
3	12 – 17	77	61.11	Moderate
4	18 – 23	23	18.25	High
5	24 – 29	0	0	Very High
Sum		126	100	

2. Analysis Prerequisites Test

a) Data Normality Test

By using a significance level of 0.01, the data is said to be normally distributed if the critical ratio (cr) of kurtosis is between ± 2.58 . Based on the results of the output data normality test on Assessment of normality, normality test results obtained by the value of the multivariate cr kurtosis $-2.58 < 2.080 < +2.58$ which means multivariate distribution is normal.

b) Linearity Test

Based on the results of linearity test the influences of X1 with Y2 have $F_{cal}=0.71 < F_{table}=1.54$. Influence X2 with Y2 have $F_{cal}=1.38 < F_{table}=1.55$ and influence Y1 with Y2 have $F_{cal}=0.83 < F_{table}=1.57$ So it can be concluded that the influence of variable class climate with physics learning result, the influence self-concept with physics learning result and achievement motivation with physics learning result are linear and are eligible for further analysis.

c) Multicollinearity Test

Table 9. The Result of Multicollinearity Test

Model		Collinearity Statistics	
		Tolerance	VIF
	Achievement motivation	.704	1.421
	Class Climate	.664	1.506
	Self Concept	.672	1.489

Dependent Variable: Physics Learning Result (Y₂)

3. Factor Analysis

a) Factor Analysis of Latent Variables

Test the fit between the theoretical models to empirical data can be seen at the level of Goodness of Fit Statistics.

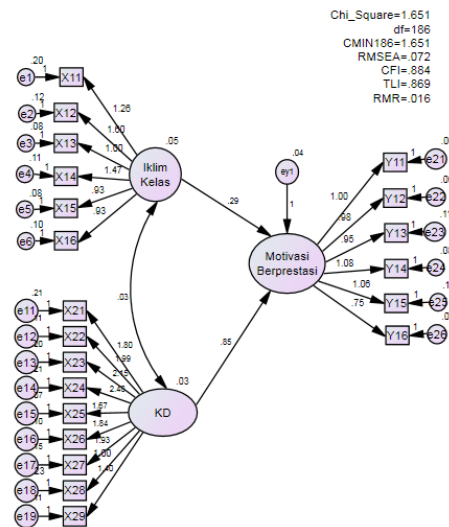


Figure 2. Initial Factor Model of Latent Variabels

The result of the factor analysis beginning shows that there are indices that do not meet the cut off value. Modifications made some errors influence of variables that have a value large change Chi-square. The modification result then re-analyzed by the results in Figure3.

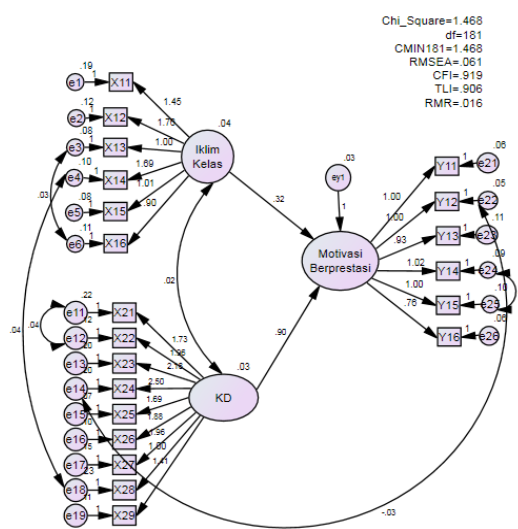


Figure 3. Final Factor Model of Latent Variabels

The final result can be seen that all the indexes have met the criteria so that these models can be received and analyzed further. Standardized regression weights can also show the influence of latent variables with the indicators.

Table 10. Bobot Regression Weights Of Latent Variables

			Estimate	S.E.	C.R.	P
X13	<---	Class Climate	1.000			
X12	<---	Class Climate	1.802	.312	5.774	***
Y11	<---	Achievement Motivation	1.000			
Y12	<---	Achievement Motivation	.964	.114	8.425	***
Y13	<---	Achievement Motivation	.921	.136	6.793	***
Y14	<---	Achievement Motivation	1.041	.134	7.759	***
Y15	<---	Achievement Motivation	1.000	.140	7.141	***
X14	<---	Class Climate	1.712	.285	6.002	***
X11	<---	Class Climate	1.462	.280	5.221	***
X15	<---	Class Climate	.905	.193	4.688	***
X16	<---	Class Climate	.863	.159	5.417	***
Y16	<---	Achievement Motivation	.741	.105	7.082	***
X28	<---	Self-Concept	1.000			
X27	<---	Self-Concept	1.753	.434	4.043	***
X26	<---	Self-Concept	1.704	.386	4.418	***
X25	<---	Self-Concept	1.499	.357	4.196	***
X24	<---	Self-Concept	2.223	.552	4.026	***
X23	<---	Self-Concept	2.030	.505	4.021	***
X21	<---	Self-Concept	1.615	.424	3.810	***
X29	<---	Self-Concept	1.268	.323	3.925	***
X22	<---	Self-Concept	1.740	.446	3.902	***

4. Model Verification and Final Model Development

Verified the theoretical model developed based on empirical data. An analysis of the picture 3 is the structural equation model of initial stages.

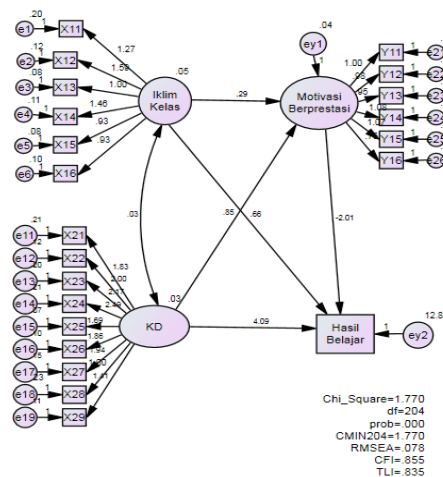


Figure 4. Structural Equation Model of Initial Stages

The results of the analysis of the initial stages in figure 4 shows that there are indices that do not meet the cut off value. Modifications made some errors influence of variables that have a value large change Chi-square. The modification result then re-analyzed by the results in Figure 5.

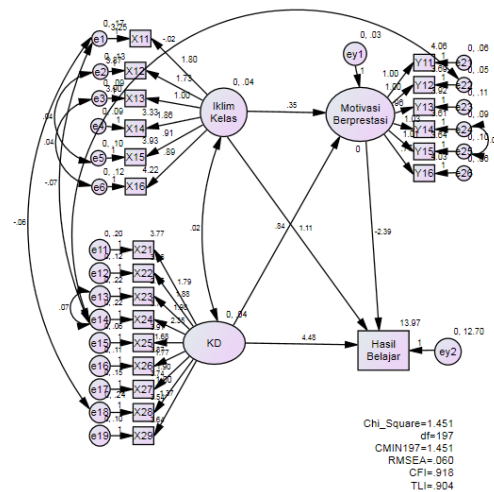


Figure 5. Structural Equation Model of Final Stages

In the final result can be seen that all the indexes have met the criteria so that these models can be received and analyzed further. Parameter of regression weighted shown in the table 11 below.

Table 11. Regression Weight Of Final Model

		Estimate	S.E.	C.R.	P
Achievement Motivation <---	Self Concept	.844	.268	3.149	.002
Achievement Motivation <---	Class Climate	.346	.173	2.006	.045
Physics Learning Result <---	Self Concept	4.481	3.483	1.286	.198
Physics Learning Result <---	Achievement Motivation	-2.385	2.212	-1.078	.281
Physics Learning Result <---	Class Climate	1.111	2.682	.414	.679

B. Discussion

1. The Direct Influence of Class Climate towards Physics Learning Result.

For the first hypothesis testing showed that the influence variables of class climate towards physics learning result described in the regression weights of final model with the estimate results $\gamma_{x1y2} = 1.111$ with p value = $0.679 > 0.05$. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that the class climate does not have positive direct influence and significant towards physics learning result.

a) The Direct Influence of class climate towards Achievement Motivation

For the second hypothesis testing showed that the influence variables of class climate towards achievement motivation described in the regression weights of final model with the estimate results $\gamma_{x1y1} = 0.346$ with p value = $0.045 < 0.05$. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that class climate have positive direct influence and significant towards achievement motivation.

b) The Direct Influence of Self Concept towards Physics Learning Result

For the third hypothesis testing showed that the influence variables of self concept towards physics learning result described in the regression weights of final model with the estimate results $\gamma_{x2y2} = 4.481$ with p value = $0.198 > 0.05$. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that self concept does not have positive direct influence and significant towards physics learning result.

2. The Direct Influence of Self Concept towards Achievement Motivation

For the fourth hypothesis testing showed that the influence variables of self concept towards achievement motivation described in the regression weights of final model with the estimate results $\gamma_{x2y1} = 0.844$ with p value = $0.002 < 0.05$. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that self concept has positive direct influence and significant towards achievement motivation.

3. The Direct Influence of Achievement Motivation towards Physics Learning Result

For the fifth hypothesis testing showed that the influence variables of achievement motivation towards physics learning result described in the regression weights of final model with the estimate results $\beta_{y_1y_2} = -2.385$ with p value = 0.281 > 0.05. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that achievement motivation does not have positive direct influence and significant towards physics learning result.

IV. CONCLUSION AND SUGGESTION

The results showed that: 1) the class climate does not have positive direct influence and significant towards physics learning result; 2) class climate has positive direct influence and significant towards achievement motivation; 3) the self-concept does not have positive direct and significant towards physics learning result; 4) self-concept has dpositive direct influence and significant towards achievement motivation, and 5) the achievement motivation has positive direct influence and significant towards physics learning result.

Based on the research results obtained, it is advisable that matters; 1)To the schools to be more looking for in-depth information about the factors that affect physics learning result students so they can be followed up quickly in order to improve learning result of students; 2)To the researcher and observer of education, based on research findings, especially some of the research hypotheses are inconsistent with the facts in theory then presumably there is further research related to the research that explore patterns of relationships between variables and look for indirect relations between the latent variables to learning result

REFERENCES

- [1] Desmita. (2010). *Psikologi Perkembangan Peserta Didik*. Bandung: PT Remaja Rosdakarya Offset.
- [2] Hadinata, P. (2009). *Iklim Kelas dan Motivasi Belajar Siswa*. Jurnal psikologi. 3(1).

ASSESSMENT COGNITIVE FOR PHYSIC: DEVELOPMENT OF MISCONCEPTION PHYSIC TEST FOR JUNIOR HIGH SCHOOL IN BANGKA BARAT WITH POLITOMOUS MODEL (PCM)

Sikto Widi Asta, Dedek Andrian
siktowidi@gmail.com, dedek.andri321@gmail.com

Abstract - This research aims to develop an instrument that measures the error concepts of physics for junior high school students in West Bangka. This research through three stages: the initial development of instruments, test, and measurement. The early development stages of instruments includes the preparation, review and validation: grating instrument, grains and assessment guidelines. Validity conducted by experienced teachers in West Bangka. The instrument has been validated been tested on 200 students. Implementation of measurement transactions are carried out on 600 students. The measurement results in the form of data politomus four categories analyzed followed Partial Credit Model (PCM) using Quest Program and Parcale. Program quest to test the suitability of the model and the level of difficulty, whereas parscale program to gain the ability, item characteristic curve, function information and SEM. The results of the study, namely: (1) a misconception physics SMP consists of 30 pieces of grains, (2) a misconception physics SMP empirically fit to PCM, (3) Whole grains instrument misconceptions physics smp in a range of criteria both for the level of difficulty is in the range from -2.00 to 2.00, (4) reliability instruments misconceptions SMP physics including high category (keofisien reliability > 0.90), so that the instrument qualifies measuring instrument physics misconceptions SMP, (5) Based on the information function, the instrument misconceptions SMP very precise physics used to check students' misconceptions capable of -2.6 to 2.00, and (6) Instruments misconceptions SMP physics can be used to measure student misconceptions smp according to PCM based on data politomus.

Keywords: *test development, misconceptions physics, and PCM.*

I. PRELIMINARY

The success of the students in learning physics is inseparable from the role of physics teachers in conveying the science of physics. Physics teacher should be able to use a good learning method possible. Physics teacher should be able to use the methods of teaching. Physics teacher must master the material and concepts taught to students.

Physics teacher must possess and understand the concept of physics well. If you do not have a physics teacher and understand the concepts of physics properly it will result in a fatal error. A fatal error in question is give and teach concepts of physics is wrong or could be mentioned as misconceptions.

There are many experts who mendevinisikan misconceptions, in this study, researchers using devinisi misconceptions of that expressed by Novak and Suparno. Novak (1984: 20) defines misconceptions as an interpretation of the concepts in a statement that can not be accepted. Suparno (1998: 95) considers misconceptions as an inaccurate understanding of the concept, the use of the wrong concept, classification examples wrong, the chaos of the different concepts and hierarchical relationship concepts that are not true. From the above misconceptions can be interpreted as a conception which is not in accordance with the scientific understanding or definition accepted by scientists.

The misconception is defined as the conceptions of students who do not fit in with the conception of the scientists, can only be accepted in certain cases and not apply to other cases and can not be generalized. The general conception is built on common sense (common sense) or intuitively constructed in an attempt to give meaning to the world they experience everyday and only a pragmatic explanation to the world of reality. Misconceptions students may also be obtained through a learning process on the level of previous education (Sadia, 1996: 13).

These misconceptions inhibit the process of acceptance and assimilation of knowledge in students for use in the learning process further (since clamps, 1988: 7). Misconceptions problem is a big problem in the teaching of physics that can not be tolerated. To identify misconceptions have been done, but until now there is still a difficulty to distinguish between students who have misconceptions by who did not know the concept. One alternative that is used to identify the misconception that developing test instruments politomus misconceptions physics model. Model instruments developed a diagnostic test that is two tier.

In 1988, Treagust tried to construct a measuring instrument that can be used to diagnose the understanding of the students. The measuring tool that is developed is a multiple choice test two levels (two-tier multiple choice). In addition to multiple-choice test two levels, has also developed a more useful diagnostic test to identify student misconceptions, namely, the concept map by Novak in 1996 and interviews by Carr in 1996 (Tuysuz, 2009). Compared with other methods, the method of two-level multiple-choice tests much easier because it is easier for the teacher in scoring compared to the other methods, making it more useful for teachers in the classroom (Tan et al, 1999).

Evaluation of test results of students based on levels which can be completed by students. Although only completed the initial phase only, examinees are getting value. The highest value of course obtained when the examinee has completed all phases of the exam in that clause. The assessment procedure is actually similar to how an individual responds to the particulars in the psychological scale. For example, an item that provides four categories of responses of "never", "rarely", "often" and "always" is analogous to the stage of completion. Resolve the matter only to the first stage analogous to the category of "never" but when it comes to the final stage, analogous to the category of "always". This assumption was later developed into a partial credit model (PCM). When it is assumed that an item follow the pattern of partial credit the individual's ability higher is expected to have a higher score than the individual who has the ability to lower (Widhiarso, 2010, p.6). According to Wright & Masters, PCM is also appropriate to analyze the response to the measurement of critical thinking and conceptual understanding in science (Van der Linden & Hambleton, 1997, pp. 101-102).

PCM was developed to analyze the test items that require several steps. PCM can be given to measures which can be undertaken by individuals. Thus, PCM is suitable to be worn on achievement tests, including the matter of physics that require problem identification to final solution.

PCM is the development of model parameters of logistic IRT 1 (1-PL) and family including Rasch model. Simple dichotomous model in Rasch model is a special case of PCM. Dichotomous model and PCM can be mixed in a single analysis (Wu and Adams, 2007). PCM is the development of items dichotomous Rasch model is applied to the item politomus. Dichotomous Rasch model item that contains only one item location parameter (degree of difficulty) and then developed with elaborate lokasibutir into several categories. PCM assumptions on which each item has the same power difference. PCM has some similarities with the Graded Response Model (GRM) on items that are scored in the category hierarchy, but the index of difficulty in each step is not necessary sequence, a step can be more difficult than the next step. This means, PCM is the development of a dichotomous Rasch model with one parameter logistic politomus the level of difficulty.

Score on the PCM category shows the number of steps to complete the details correctly. Score higher categories showing greater ability scores lower category. In PCM, if an item has two categories, the probability equation to equation Rasch model, as the equations specified by Hambleton and Swaminathan (1985). Therefore, PCM can be applied to grain politomus and dichotomous.

The survey shows that the scoring of multiple-choice tests using dichotomous model, meaning that if the item is given a score of 1 and if one is given a score 0. Scoring is not using politomus fairer model for considering measures test solution. This dichotomous scoring model

does not appreciate stages of completion, because with a different error level get the same score of 0., thus scoring model is certainly unfair.

Based on the researchers are interested in developing the test misconception physics. Through this instrument is expected to facilitate teacher evaluation to student learning.

II. RESEARCH METHODS

This research was the development of a quantitative approach. This study uses a model oriondo wilson and the model that has been modified.

The study began in January 2015 until December 2015. The initial development tests in the form of test preparation, validation, and test assembly was conducted in May 2015 to September 2015. The trial took place in September 2015. The study was conducted in secondary schools in the West Pacific region.

The subject of research students in the district smp western farts. Number of research subjects as many as 600 participants.

Step-by-step development of instruments such as tests using a modified Model and Model Oriondo Wilson and Antonio, namely: (1) the design of the test, (2) the trial tests, and (3) test assembly. The design phase of the test include: (1) determining the test objectives, (2) the determination of competence tested, (3) the determination of the material tested, (4) preparation of grating tests, (5) that the item is based on the principles of development tests, (6) validation test items, (7) repair and assembly of test items, and (8) establishing guidelines for scoring.

The test phase tests include: (1) the determination of the trial subjects (SMA), (2) the implementation of the trial, and (3) data analysis test results. The last stage in the development of this test is a test assembly. Stages of development of these tests are presented in Figure 1.

Relating the sample size, for some expert analysis IRT measurements should be 200 to 1000 people (Seon, 2009, p.3). For Rasch analysis, the samples used between 30 to 300 people (Bond and Fox, 2007, p. 43; Keeves & Masters, 1999, pp.12-13). Reckase (2000) concluded that the minimum sample size was good for Estimating the three parameters that include: discrimination (different power), the level of difficulty, and pseudoguessing is 300 (Haladyna, 2004, p.206). So, with the model PCM 1PL learners who were subjected try as many as 500 people are already more than enough.

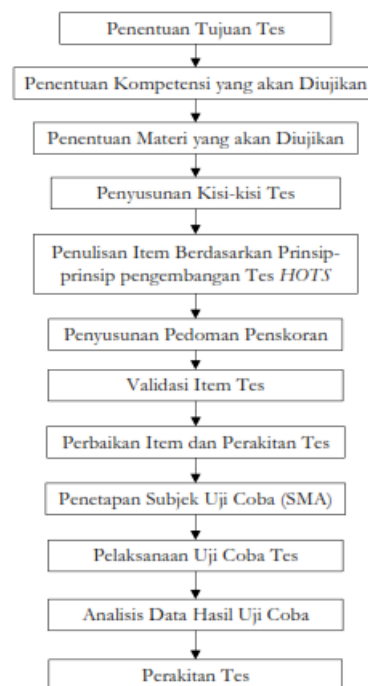


Figure 1. Step-by-step development of its instrument of accession.

This study data analysis using 1 PL Partial Credit Model (PCM 1PL) for fit testing of test items for the subjects of Physics misconceptions smp. The basic consideration is used, the first is that PCM as an extension of the Rasch model is a model 1-PL, can use the sample which is not as if performing the calibration data is politomus using model 2-PL or 3-PL (Keeves & Masters, 1999, pp.12 -13). Second, that the response characteristics of the test items for the subjects Physics misconceptions smp follow PCM. Data analysis was performed on several aspects, namely: (1) the suitability item instrument, (2) reliability, (3) the item characteristic curve (ICC), (4) the index of difficulty, and (5) the function of information and SEM.

III. RESULTS AND DISCUSSION

A. Test Assumptions Unidimensional

To test the assumption unidimensional factor analysis. Before the factor analysis was performed testing the feasibility of using test analysis KMO-MSA and Bartlett Test. According to Anderson (1998: 88) is the factor analysis requirements Keiser-Meiyer Olkin value greater than 0.5 and significant values Bartlett Test of less than 0.05. KMO-MSA test results and Bartlett Test can be seen in Table 1 below:

Table 1. Test results KMO-MSA and Bartlett Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.987
Approx. Chi-Square		11420.670
Bartlett's Test of Sphericity	df	435
	Sig.	.000

Based on the test results KMO-MSA and Bartlett Test as set forth in Table 1 it is concluded that the KMO-MSA test results and test Bartlett's test misconceptions physics SMP qualify namely: Keiser-Meiyer Olkin value greater than 0.05 and significant value Test Bartlett is less than 0.05.

To get items that measure the same dimension extraction process is carried out so that the resulting number of factors. Each factor formed are eigenvalues maintained above 1.00 (Norusis, Marija: 155-161). Note the picture 2 below:

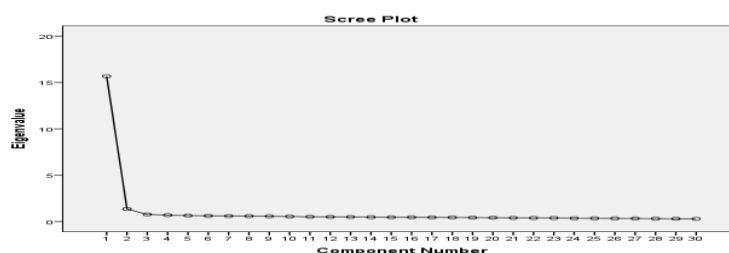


Figure 2. Value Eigenvalue

From the results of scree plot and eigenvalues value it can be concluded that the items were used in the tests misconceptions physics smp measure the same dimension.

B. TEST FIT MODEL

Testing goodness of fit for the test as a whole as well as grains with Quest program. Testing determination fit every item on the model following the rules of Adam & Khoo (1996: 30), ie all the items fit to the model if the value INFIT MNSQ between 0.77 to 1.30. With the acceptance of grain boundaries using INFIT MNSQ or fit according to the model (between 0.77 up to 1.30) and using INFIR t degan limits -2.00 to 2.00, the obtained item that meets the goodness of fit. The test results fit model using the Quest program can be seen in the following figure:

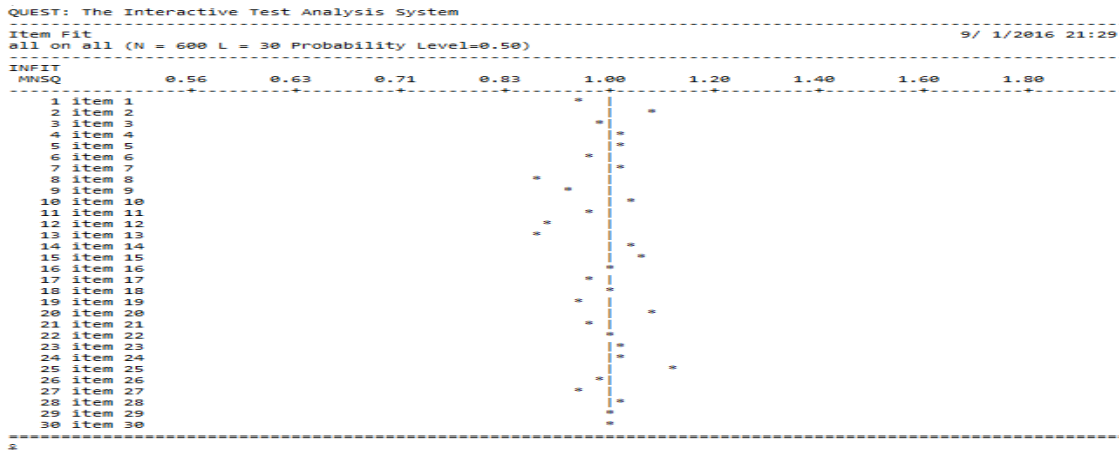


Figure 3. INFIT MNSQ on misconceptions physics test SMP

Quest output based on the results it can be concluded that the test misconceptions physics smp PCM fit to the model it can be seen from INFIT MNSQ value of between 0.77 to 1.30.

C. ESTIMATED RESULTS

Based on data analysis, obtained characteristics test instrument misconceptions SMP physics. Characteristics of test instruments misconceptions SMP physics can be seen in table 2 below:

Table 2. Characteristics of test instruments misconceptions physics SMP

No	Uraian	Estimasi untuk butir	Estimasi untuk testi
1	Nilai rata-rata dan simpangan baku	$0,00 \pm 1,24$	$0,41 \pm 1,79$
2	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	$0,00 \pm 1,22$	$0,41 \pm 1,77$
3	Reliabilitas	0,97	
4	Nilai rata-rata dan simpangan baku INFIT MNSQ	$0,99 \pm 0,06$	$0,99 \pm 0,27$
5	Nilai rata-rata dan simpangan baku OUTFIT MNSQ	$0,99 \pm 0,09$	$0,99 \pm 0,36$
6	Nilai rata-rata dan simpangan baku INFIT t	$-0,09 \pm 0,95$	$-0,03 \pm 0,98$
7	Nilai rata-rata dan simpangan baku OUTFIT t	$-0,01 \pm 0,85$	$0,05 \pm 0,76$
8	Butir atau testi sekor 0	0	0
9	Butir atau testi sekor perfect	0	0

D. DIFFICULTY LEVEL

The results of the index level of difficulty can be seen in Figure 4 below:

Item Estimates (Difficulty and Taus) In input Order									
all on all (N = 600 L = 30 Probability Level=0.50)									
ITEM NAME	SCORE	MAXSCR	DIFFCLTY	TAU/S	1	2	3	4	
1 item 1	1412	1800	-1.64	-1.09	-0.64	1.73			
			.07	.25	.16	.11			
2 item 2	473	1800	2.04	-1.86	0.86	1.00			
			.07	.11	.15	.21			
3 item 3	1220	1800	-0.93	-1.97	0.58	1.39			
			.06	.19	.12	.11			
25 item 25	818	1800	0.64	-2.72	0.67	2.06			
			.07	.14	.11	.16			
26 item 26	1236	1800	-0.85	-1.29	-0.27	1.56			
			.06	.18	.13	.11			
27 item 27	1110	1800	-0.27	-0.41	-0.62	1.03			
			.06	.14	.13	.11			
28 item 28	1159	1800	-0.68	-2.27	-1.33	3.60			
			.08	.24	.13	.14			
29 item 29	1356	1800	-1.39	-1.39	-0.13	1.53			
			.07	.22	.14	.11			
30 item 30	1241	1800	-1.01	-1.61	-1.27	2.89			
			.08	.23	.14	.12			
Mean			0.00						
SD			1.23						

Figure 4. Index of difficulty

Based on the index of difficulty so the easiest is no 8 and the most difficult is the question no 11.

E. CHARACTERISTICS OF ABILITY STUDENTS

Distribution ability learners SMP in western part expressed in figure 5 below:

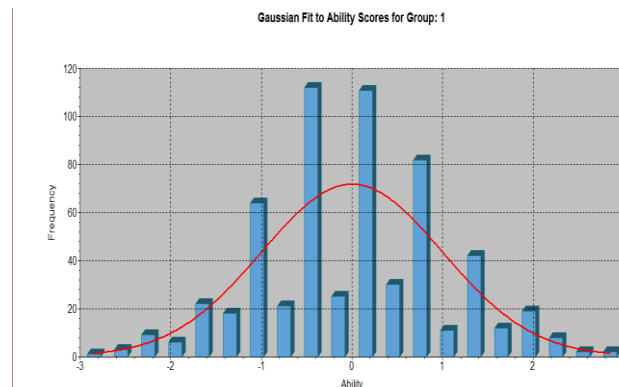


Figure 5. Graph of the distribution of the ability of learners SMP in western Bangka.

Based on the figure 5 it can be said that the students' ability lowest is 3.00 -3.00 and the highest ability. Conclusion on the student's ability: the ability -3.00 can mean that students did not understand the physics even experienced misconceptions, and the ability of students 3.00 can mean that students really understand all of physics untuk both theories, concepts and formulas properly.

F. ICC (ITEM CHARACTERISTIC CURVE) AND IIC (ITEM INFORMATION CURVE)

ICC and IIC dilihat by enabling run-plot on the menu to open a chart Parscale IRT. In PCM calibration, to the 30 items showed a matrix plot ICC seperti figure 6 below:

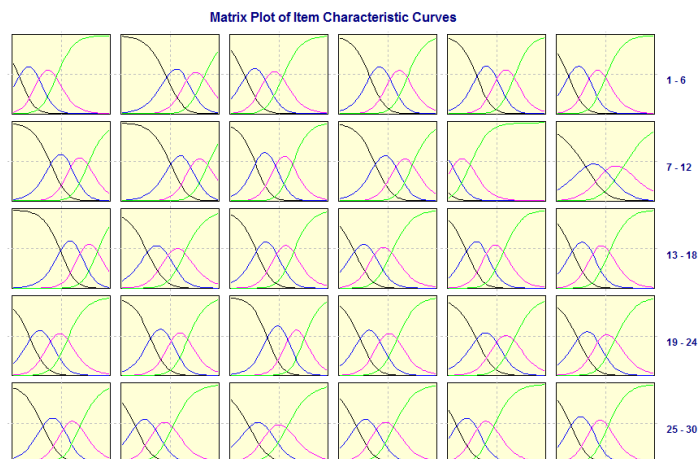


Figure 6. ICC test misconceptions physics SMP

G. APPROPRIATE SKILLS OF INSTRUMENTS FOR ACTIVITY MEASUREMENT

According to analysis by Parscale obtained information and SEM function graph presented in Figure 7 below:

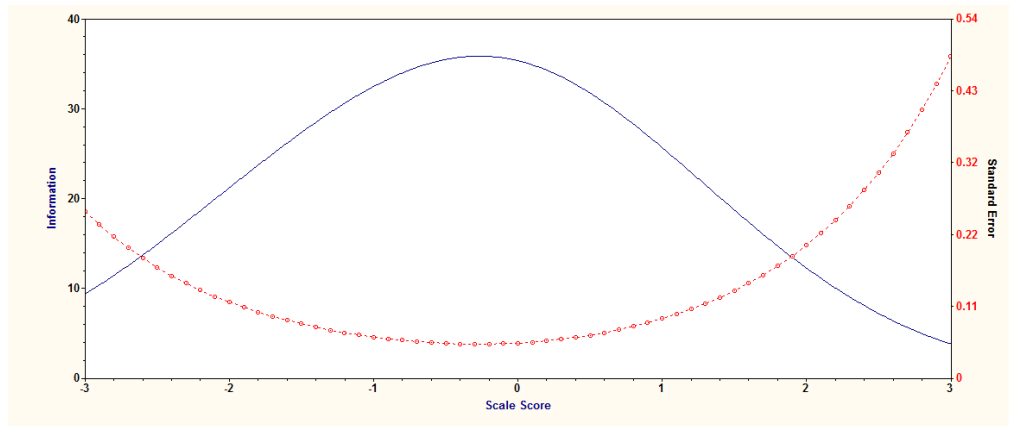


Figure 7. Functions Information and SEM

Based on the figure 7 tests misconceptions physics SMP accordance with learners who have the ability: $-2.6 \leq \theta \leq 2$

IV. CONCLUSION

The results of the study, namely: (1) a misconception physics SMP consists of 30 pieces of grains, (2) a misconception physics SMP empirically fit to PCM, (3) Whole grains instrument misconceptions physics SMP in a range of criteria both for the level of difficulty is in the range from -2.00 to 2.00, (4) reliability instruments misconceptions SMP physics including high category (reliability coefficient > 0.90), so that the instrument qualifies measuring instrument physics misconceptions SMP, (5) Based on the information function, the instrument misconceptions SMP very precise physics used to check students' misconceptions capable of -2.6 to 2.00, and (6) Instruments misconceptions SMP physics can be used to measure student misconceptions SMP according to PCM based on data politomus.

REFERENCE

- [1] Adams, R. J. & Khoo, S. T. (1996). *Quest: The interactive test analysis system version 2.1*. Victoria: The Australian Council for Educational Research.
- [2] Bond, TG and Fox, CM. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences (2nd edition)*. Mahwah, NJ: Lawrence Erlbaum. Bonk.
- [3] Hambleton & Swaminathan. (1985). *Item Response Theory Principles and Applications*. Boston: Kluwer Nijhoff Publishing.
- [4] Haladyna, T. M. (2004). *Devoping and Validating Multiple Choise Test Items*. New Jersey: Lawrence Erlbaum Associates, Inc.
- [5] Keeves, J. P. & Alagumalai. (1999). New Approach to measurement. Dalam: Masters, G.N. & Keeves, J.P. (eds.). *Advances in Measurement in Educational Research and Assessment* (pp.23-42). Amsterdam: Pergamon, An imprint of Elsevier Science
- [6] Keeves, J.P. & Masters, G.N. (1999). Introduction. Dalam: Masters, G.N. & Keeves, J.P. (eds.). *Advances in Measurement in Educational Research and Assessment* (pp.1-22). Amsterdam: Pergamon, An imprint of Elsevier Science
- [7] Novak, J.D and Bob Gowin. 1985. *Learning How to Learn*. Cambridge University Press.
- [8] Paul Suparno. 2013. Miskonsepsi Dan Perubahan Konsep dalam Pendidikan Fisika. PT grasindo Jakarta.
- [9] Sadia. 1996. Pengembangan Model Belajar Konstruktivis dalam Pembelajaran IPA di SMP. (Suatu Studi Eksperimental dalam Pembelajaran Konsep Energi Usaha dan Suhu di SMPN I Singaraja). *Disertasi* (tidak diterbitkan). IKIP Bandung.
- [10] Seon, Hi Sin. (2009). How to tread omitted respons in Rasch model based equating [Versi elektronik] *Practical Assessment, Research&Evaluation*. ISSN15317714, Volume 14, Number 1, p: 1-8
- [11] Van der Linden, W. J & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: SpringerVerlag New York, Inc
- [12] Widhiarso, Wahyu. (2010). Model politomi dalam teori respons butir. Yogyakarta: Psikologi UGM
- [13] Wu, M., & Adams, R. (2007). *Applying the Rasch model to psychosocial measurement: A practical approach*. Melbourne: Educational Measurement Solutions.

Identifying of Undergraduate's Analytical Ability About Electric Current In Transistor Using Isomorphic Assesment

Sri Hartini¹, Dewi Dewantara², Misbah³, Syubhan Annur⁴

^{1,2,3,4} Department of Mathematics And Natural Science, Lambung Mangkurat University
t2n_fis@yahoo.com

Abstract - One objective of the courses is the student basic electronics, student is able to analyze the working principle of a transistor. Therefore, need research to analyze the analytical ability of students about the current in the transistor. Analytical ability measured in this study is the ability matching, specifying, generalizing and analyzing errors. The instruments used are isomorphic assessment. The subjects were students of the University of Physical Education Mangkurat which follows the course Basic Electronics. Based on the analysis we concluded that 90% had less good analytical ability of the current on the transistor.

Keywords: transistor, isomorphic assessment, analytical skills

I. INTRODUCTION

Transistor components widely used among others as a current amplifier, generating vibrations, connect and disconnect the electrical signal and transform alternating current into direct current stable [1]. Transistor is an active component made of semiconductor materials. Semiconductors consists of two types namely p-type and n-type [2]. Transistor is a solid version of the triode vacuum tube. Based on the structure, the transistor can be divided into the n-p-n transistor and the p-n-p transistor. N-p-n transistor is composed of layers of the type p sandwiched between two layers of type n. P-n-p transistor is a transistor consisting of the n type layer inserted between two layers of p [3].

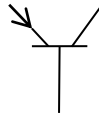


Figure 1. P-n-p transistor

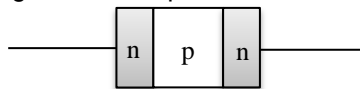


Figure 2. N-p-n transistor

The middle section is called the transistor base, while both ends are the emitter and collector. Arrows in the diagram, the base is in the middle and the line next to the arrow that flies is emitter. Lines that do not have an arrow collector. Current will flow from the emitter voltage source through to get to the base. An arrow on the emitter will show the poles of the transistor. Arrow sign to indicate that the base has a p-type emitter, while the arrow out of the base shows that the emitter has a kind of n.

In this type of transistor n-p-n and p-n-p currents emitter, base and collector is positive if current flows into the transistor. The voltage drop from the emitter to the base, the collector to the base, and the collector to emitter is positive if the first positive terminal. Under normal circumstances, because the relationship is supplied forward base emitter, the emitter current of transistor negative for n-p-n transistor and positive for p-n-p to "benchmark" is selected. If the collector-base relationship fed backward, the base-collector voltage of the transistor is positive for n-p-n transistor and negative for p-n-p transistor [3].

One objective of the course Basic Electronics is that students are able to analyze the workings of the transistor. One of analytical abilities is to analyze the current at the base, emitter, and collector. Analysis capabilities are part Bloom's Taxonomy that the cognitive level four [4]. This ability is related to the ability for deliver material into the portions smaller, so that the organizational structure can be understood. Analytical capabilities include: matching, classifying, analysis errors, generalizing, and specifying [5]. Matching is an activity to identify similarities and differences as well as the relationship between the components of the material problems in physics. Classifying is an activity identifies the superordinate and subordinate categories where knowledge of physics related to the problem can be arranged. Analyzing errors are analyzing logic, fairness, and accuracy based on knowledge of physics. Generalizing an activity to build new generalizations or principles of available knowledge of physics. Specifying an ability to evoke latest applications or logical consequence of the physics knowledge available.

One way that can be used to determine the analytical ability of students is to use the assessment isomorphic. Assessment isomorphic capable of identifying students analytical ability and get the right information about the students' ability to analyze current on the transistor. When students have to analyze the current and voltage of the transistors work, then students must be thorough and precise work on the problems isomorphic. This suggests that students should be careful in looking at the concept of equality between several issues which have different features (isomorphic) [6]. Thus students are able to identify similarities and differences as well as the relationship between the components of the problem, analyze logic, fairness and accuracy, building new generalizations, and raise the ability of the latest applications.

Reference [7] have used the problems isomorphic to see the different contexts in students' understanding of the concept. Reference [6], [8], [9] stated that when given about the same concept, even though it looks the same matter by experts, the question may look very different by the students. Based on previous studies of the isomorphic, then the researchers intend to use about isomorphic to identify students' analytical skills related to the base current, the emitter and collector of the transistors work. This study aims to identify the students' analytical skills on transistor material based tests isomorphic

II. METHODS

This research subjects were students of Physical Education on Lambung Mangkurat University at second semester of 2015/2016 academic year which follows the course Basic Electronics. The instrument consists of 10 questions tests isomorphic that include of 5 questions about p-n-p transistor and 5 questions about n-p-n transistor. The data were analyzed by descriptive quantitative and qualitative.

III. RESULT AND DISCUSSION

Table 1 show that percentage of students' analytical ability. Table 2 show that distribution answers of isomorphic assessment.

Table 1. Percentage of students' analytical ability

No	Category analysis capabilities	Percentage
1	Excellent	0.00
2	Good	0.00
3	Average	10.0
4	Less	90.0
5	Weak	0.00

Table 2. Distribution Answers of Isomorphic Assessment

Category	Number of Problem	Percentage
Students have really gives the reasons for the currents at the emitter-base-collector transistor n-p-n	1	87.5
	2	2.50
	3	85.0
	4	5.00
	5	5.00
Students have really gives the reasons for the currents at the emitter-base-collector transistor p-n-p	6	0.00
	7	55.0
	8	0.00
	9	90.0
	10	2.50

Problems are used in this study is a matter of isomorphic. Problems 1, 2, 3, 4, and 5 isomorphic about the current and voltage at the n-p-n transistor. Problems 6, 7, 8, 9, and 10 isomorphic about the current and voltage at the transistor p-n-p. These questions are using the same concept but in a different form. The series contained in the question to digging the analytical ability of students for determining the direction of the base current, collector, and emitter. Transistors can be illustrated with arrows diagram form as Figure 1 and may also like Figure 2. Flows on the transistor to depend the voltage at the transistor supply there is the power supply forward and backward supply. Each problem is given in a variety of different images. Based on observations of 10% of the students have average analytical ability, and the remaining 90% who had less analytical ability of the current on the transistor.

On problem 1, 87.5% of the students have been really gives the reasons for the currents at the emitter-base-collector. This suggests that in analyzing the transistor n-p-n, the majority of students have been able to identify the correlation between the components of the problem, and to analyze the logic, fairness, and accuracy based on knowledge of basic electronics.

On problem 2, 2.50% of the students have been really gives the reasons for the currents at the emitter-base-collector n-p-n transistor. Whereas the given problem has a concept and a voltage source similar with problem 1 but on the problem 1 where is transistor illustrated with arrows diagram like Figure 1 and problem 2 is described as Figure 2. This shows that most students still unable to build new generalizations or concept principles of the currents at the emitter-base-collector.

On problem 3, 85.0% of the students have been really gives the reasons for the currents at the emitter-base-collector. Problem 3 has the form of a picture that similar with the picture of problem 1 but there are given the constraints on the part of the emitter, base and collector. In this problem where the majority of students have been able to identify the correlation between the components of the problem; analyze logic, fairness, and accuracy based on knowledge of basic electronics; as well as building new generalizations or principles of the concept of the currents at the emitter-base-collector.

On problems 4 and 5, 5.00% of the students have been really gives the reasons for the currents at the emitter-base-collector on the n-p-n. 5 pieces of the problem of the transistor n-p-n, 3 pieces already disadvantaged students well analyzed. This is because students are only given the picture that has been described in detail by the teacher. When the matter is modified in different features, students difficult to see the similarity of concepts contained in the question.

On problems 6, 7, 8, and 10 indicate that most students have not been able to identify the correlation between the components of the problem; analyze logic, fairness, and accuracy based on knowledge of basic electronics; as well as building new generalizations about the currents at the emitter-base-collector. These questions requires analysis capabilities be based on the concept of transistor p-n-p is in the form of different problems. The same reasons as the failure of students in analyzing problems n-p-n transistor, when given about the same concept, the questions look very different by the students. That was causes students to think of other concepts and other ways to solve the problem and do not aware that the solution is the same as the questions about the transistor p-n-p other. This is consistent with previous studies that when given about the same concept, although the question looks just by experts, the question could be seen very differently by students [6], [8], [9].

When the student answered correctly in one of 5 problems with the same concept, then right again in answering the next problem, then the ability of matching, classifying, analyzing errors, generalizing, and specifying excellent category. However, if the question is answered correctly fraction while most of the others are wrong, then the student needs to further improve its analysis capabilities. The cause analysis capabilities are categorized as less and weak may be caused by the analytical ability of students who are not familiar maximized. Another cause is the low mastery of concepts, misconceptions, or less understanding of the concept. Therefore was suggested for further research to identify the causes of less student in analytical skills.

IV. CONCLUSION

Based on the analysis of test results using isomorphic assessment was concluded that 90% had less analytical ability of the current on the transistor.

ACKNOWLEDGMENT

This research was supported by Faculty of Teacher Training and Education Lambung Mangkurat Lambung Mangkurat University (PNBP fund 2016). Authors thank to Dean of Faculty of Teacher Training and Education Lambung Mangkurat University for research permission.

REFERENCES

- [1] M.Suyanto. 2013. Aplikasi Sistem Inventer 1 Fasa Dengan Kapasitas Beban 1200 Watt. Jurnal Teknologi Technoscianta. Vol 6. No. 1 pp. 103-109.
- [2] E. Aditya, Suudi, & Endarko. 2012. *Transistor*. Jurnal Transistor Vol 1 No. 1 pp. 1-35
- [3] D. Chattopadhyay, P.C. Rakshit, B. Saha, N.N. Purkait. 1989. Dasar Elektronika. Terjemahan Suyanto. Penerbit Universitas Indonesia.
- [4] R.E., Teodorestu, C. Bennhold, G. Feldman, & L. Medsker, 2013. New Approach to analyzing physics problem: A Taxonomy of Introductory Physics Problems. Physical Review Special Topics- Physics Education Research, Vol 9 No. 01 pp. 1-20
- [5] D.R. Krathwohl, 2002. A Revision of Bloom's Taxonomy: An Overview. Theory into Practice, Vol 41 No. 4, pp. 212-264
- [6] L.Y, Shih & S. Chandralekha. 2013. *Using an Isomorphic problem pair to learn introductory physics: transferring from a two step problem to a three step problem*. Physical Review Special Topics- Physics Education Research. Vol 9 No. 020114 pp.1-21.
- [7] B. Pablo & Z. Genaro. 2014. *Force, velocity, and work: The effect of different contexts on students understanding of vector concepts using isomorphic problems*. Physical Review Special Topics- Physics Education Research. Vol 9 No. 020115 pp.1-15.
- [8] T.H. Michelene, P.Chi, J. Feltovich, & G.Robert. 1991. *Categorization and representation of physics problems by expert and novies*. Cognitive Science. Vol 5 no. 2 pp. 121-152.
- [9] P. Hardiman, R. Dufresne, & J. Mestre. 1989. *The relation between problem categorization and problem solving among experts dan novices*. Memory and Cognition. Vol 17 No. 5 pp. 627-638.

A Performance-Based Assessment as a Current Trend in ELT: Investigating Its Washback Effects on Secondary-School Student Learning

Sumardi

SMA Negeri 1 Sragen

arif_sumardi74@yahoo.co.id

Abstract – Assessment is an integral part of an instruction since it enables English language teachers to visualize to what extent the success of their instruction. However, they still override its functions to improve student learning. Even, most of English language teachers focus more on traditional formats of assessment, such as multiple choice, true false, matching type, etc and pay little attention to the washback effects of those kinds of assessment on students' learning. A traditional assessment has at least two fundamental flaws. There is a little chance to identify students' factual skills as it just focuses on formal assessment. In addition, it only leads the students solely to obtain high scores. This orientation can be misleading to the nature of language instruction itself. Therefore, this research article explores the washback effects of performance-based assessment as a current trend in English Language Teaching (ELT) on secondary-school student learning. To investigate this issue, in-depth interview, classroom observation, and document analysis were deployed as the data collection methods. The merits of the research are to show how to deploy performance-based assessment and to identify its washback effects on student learning. It is eventually known that performance-based assessment encouraged or motivated students to learn more. Students' perception on this type of assessment was positive and they strived to be more successful in their learning.

Keywords: *performance-based assessment, washback effects, student learning.*

I. INTRODUCTION

Language assessment is a field of study under the umbrella of applied linguistics. Its main focus is the assessment of first, second and foreign languages in the school, college, or university context. The assessment may include listening, speaking, reading, writing, or may be an integration of two or more of these skills. There are at least two modes of language assessment that English language teachers may follow to assess student learning. The first one is traditional standardized achievement tests. It is mostly in the format of paper and pen tests, such as multiple-choice, true false and matching type test. These kinds of test are used for high-stake decision in assessment and accountability system. Traditional standardized achievement tests are still widely used by English language teachers throughout Indonesia. The second mode is authentic assessment. It refers to a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills (Mueller, 2014). One of the formats of authentic assessment is performance-based assessment. Performance-based assessment can be used by English language teachers to assess student learning by following numerous methods of assessment, such as oral presentation, story-telling, dramatic reading, and English debate. It is a more challenging format of assessment to develop and implement than other formats of assessment as it is a key to student engagement. Those two kinds of assessment have their own strengths and weaknesses to assess student learning outcomes in English Language Teaching (henceforth ELT) classroom.

A traditional assessment, especially in the format of multiple-choice test, is still considered as more accessible assessment for English language teachers. Although this kind of assessment is relatively difficult to develop, it is actually easier to administer than that of performance-based assessments. When a multiple-choice test is applied to assess student learning outcomes, teachers may objectively score student learning outcomes as no teachers' subjectivity may

influence on the results of scoring. Besides, the accountability of test results can be achieved as this test format can be scored objectively by English language teachers. However, the common limitations may occur as well when multiple-choice tests are applied to measure student learning outcomes. Abedi (2010) states that traditional assessments do not afford an opportunity for students to present a comprehensive picture of what they know and are able to do in content areas. In addition, multiple-choice test as one of traditional assessment formats usually fails to assess higher order skills and other skills essential for functioning in schools or work settings (Haney & Madaus, 1989; Neill & Medina, 1989; O'Neil, 1992; Wiggins, 1989 in Pierce & O'Malley, 1992). Besides, multiple-choice tests are generally incomplete since they portray an individual at a single moment in time within particular context (McTighe & Ferrara, 2011). Due to those limitations of traditional assessment, there has been a current trend and a growing interest among English language teachers in performance-based assessment.

Performance-based assessments may help to fill the gaps that probably appear when the traditional assessment is applied for assessing student learning outcomes. A research by Bass, Magone, & Glaser (2002) identified that performance-based assessments allow all students, especially those with different language backgrounds, to engage in cognitively complex activities such as generating strategies, monitoring work, analyzing information, and applying reasoning skills. Another research by Linn & Burton (1994) revealed that performance-based assessments have appeal as assessments that are better reflections of criterion performances that are of importance outside the classroom. Related to English language assessments, performance-based assessments can help actually to identify language factors that influence assessment outcomes (Abedi, 2010). Similarly, Goldschmidt, Martinez, Niemi, & Baker (2007) found that performance-based assessments as the open-ended assessments improve the chances for English language learners (ELL learners) to engage with language production and learning, offering unique opportunities for ELL learners express their knowledge in a broader sense than the limited linguistic opportunities given to them in traditional multiple-choice items. In short, performance-based assessments may be used by English language teachers to assess the students' real language performances.

The washback effects may usually be associated with that of the application of performance-based assessments applied in ELT classroom. Washback is the effects of tests and/or assessments on teaching and learning (Doroban, 2007) and it can be positive or negative washbacks, (Cheng & Watanabe, 2004) both for students and English language teachers in ELT classroom. Having implemented performance-based assessments, teachers may try hard to find a new method of teaching to facilitate students to learn better in ELT classroom or students may be well-motivated to learn more to fulfill their needs in learning English as foreign language. In addition, a research by Darling-Hammond (2006) found that performance assessments that require students to evaluate and solve complex problems, conduct research, write extensively, and demonstrate their learning in projects, papers, and exhibitions have proven key to motivating students and attaining high levels of learning. To summarize, a well-design of performance-based assessment may motivate students to engage totally during the class as well as encourage teachers to facilitate effectively student learning in ELT classroom.

A current trend of assessing student learning outcomes tends to provide emphasis on the learning by using performance-based assessment as one of authentic assessments for assessing the students' real language performances. It has become a central issue as English language teachers are required to apply authentic assessments to assess student learning outcome. It is also recently required by curriculum 2013 that English language teachers need to shift their choices to assess student learning outcomes; not only using traditional assessments but also using performance-based assessments. That is why performance-based assessments are now getting its popularity among the English language teachers as curriculum 2013 requires them to apply such an authentic assessment as a method of educational assessment.

Realizing the importance of performance-based assessment to assess the student learning, this research article, then, tries to explore the washback effects of the application of performance-based assessment as a current trend in ELT classroom; especially, when it is applied for secondary-school students. The secondary-school students were selected as the subjects of this research as they have had relatively adequate levels of language proficiency to communicate both in oral and written English compared with those of primary-school students. For this purpose, the subsidiary research questions are formulated as follows:

1. How a performance-based assessment should be designed to assess the secondary-school student learning in ELT classroom?
2. Are there any washback effects of performance-based assessment on the secondary-school student learning?

II. RESEARCH METHOD

The nature of present study was actually a classroom action research with two parallel cycles and consists of several phases for every single cycle; such as planning, implementing, observing, and reflecting on the particular actions. This study aimed at exploring the washback effects of performance-based assessment as a current trend in English Language Teaching (ELT) on secondary-school student learning. In addition, this study focused merely on investigating the washback effects of performance-based assessment which was deployed in speaking class. Thirty two students sitting in the twelfth grade of a secondary school in Sragen, Central Java volunteered to participate in this research project. They had received formal English instruction for six years or more and their language ability ranged from elementary to intermediate levels based on the results of pre-test on their speaking skills. Besides, students' initial motivation to enroll the speaking class and to speak in English was also identified remain low.

Language assessment is an integral part of an instruction since it enables English language teachers to visualize to what extent the success of their instruction. Therefore, performance-based assessment was part of an authentic assessment deployed in English language instructions. To investigate its washback effects on the secondary-school student learning, the oral presentation as one of the methods of performance-based assessment was designed by following several stages: (1) teachers provided firstly a stimulus related to the topics to be discussed during the instructional process; (2) teachers established the students into groups consisting of three or four students in every single group; (3) each group of students was given the opportunity to identify the varieties of social issues containing aspects of the pros and cons and occurring in the local communities; (4) each group of students carried out an intensive discussion with regard to the social problems to identify why those problems provoke the pros and cons; (5) each group of students was, then, given an opportunity to deliver an oral presentation for 10 to 15 minutes; (6) as one of the groups had completed a presentation, the other groups were given the opportunity to deliver a range of questions in connection with the materials presented; and (7) each member of the group, who was doing a presentation, then provided responses to the variety of questions submitted by members of other groups. Those stages were the hierarchical activities to realize or to manifest a performance-based assessment in speaking class.

Empirical data were collected through in-depth interview, classroom observation, and document analysis. The in-depth interviews served as verbal justifications from the participants for what were observed in the classroom and for what were unknown in the classroom observation. The classroom observation data were video recorded, while interview data were audio recorded. Finally, documents related to the theories underlying this project were analyzed to provide a complete depiction on how this project should be conducted.

The empirical data were qualitatively analyzed. This qualitative analysis entails an interpretative enterprise. To discern multifaceted qualitative data, micro-interaction analysis using Anderson's (2009) interaction framework was deployed to examine the collected data garnered from classroom observations. All the data were transcribed and reviewed, and the data showing participants' actions, moves, and interactional patterns were tabulated. The data were categorized into moment-by-moment interactions (teacher-student and student-student interactions), characterizations of such interactions (how much both the teachers and the students valued such interactions), and interactional patterns (how students interacted with their peers). These three layers of the analysis allowed the author to capture some emerging findings relevant to the two research questions.

III. RESEACH FINDINGS AND DISCUSSION

Oral presentation as one of the methods related to performance-based assessment was designed for assessing the students' speaking skills. Oral presentation was carried out through several stages as they were stated previously in this article and the students had to perform an oral presentation in groups. Thus, group-based oral presentation underlined how performance-based assessment should be carried out during this research project. The students' initial

motivation to speak in English before this kind of assessment was carried out would be firstly outlined in this article. Furthermore, the washback effects of performance-based assessments on the student learning would be outlined in the next section of this article.

A. Description of Students' Initial Motivation to Speak in English

The students' initial motivation to speak in English for most students was identified still very low. Generally, the students were afraid of making any mistakes to speak in English so that their motivation and self-confidence to speak was exceptionally low. Such a condition has resulted in their speaking ability was ultimately inadequate. The students' motivation to speak and their ability to speak in English were the two interrelated-variables. The higher the students' motivation, the better the students' English proficiency. Similarly, Aidogmus, Aksu & Rich (2014) state that the motivation is very closely linked to the students' ability to communicate orally. In addition, Quadir (2014) also points out that the motivation has a strong correlation with verbal communication skills (oral communication), especially for students who are learning English as a foreign language. The latter opinions explicitly confirmed that a teacher should be able to design an appropriate method of assessment in order to motivate the students to learn better and to communicate well and spontaneously in English. This is a very important issue because students' ability to communicate orally and spontaneously is a major indicator of the students' success in learning English.

Based on initial interviews conducted by the researchers, there were several reasons why the students' motivation to communicate orally was still exceptionally low. Among those reasons were (1) the students were afraid of making any mistakes when they wanted to communicate orally in English; (2) the students' mastery on English vocabularies was still very limited; and (3) the students' level of confidence to speak in English was still very low as well. Another dominant factor leading to the students' low motivation was that a method of assessment used by teachers tended to be monotonous and mechanistic. It did not challenge the students to learn actively during the speaking class in order to achieve the desirable learning outcomes. When the assessment of speaking was going on, teachers just asked the students to set up a dialogue with their friends in group and then practiced it in front of the class.

This activity took place repeatedly when speaking class was carried out in speaking class, so that the students were not challenged to speak in English spontaneously. Visually, the students seemed to be able to speak in English fluently, but the students' fluency in speaking English was because the students had first memorized the utterances they prepared before practicing a dialog in front of the class. Thus, the students were generally not able to communicate orally and spontaneously. In addition, the teacher did not provide a proportional opportunity for the students to speak in English as well. That was why the teacher (the researcher) needed to cope with those problems by implementing a particular treatment in order to motivate the students to learn more and to allow them to be more active during the speaking class. A treatment provided by the researcher was actually a performance-based assessment which was implemented in two parallel cycles. Although a performance-based assessment was implemented in two parallel cycles (cycle I and cycle II), it was slightly modified in cycle II so that there was different effects on student learning compared with that of performance-based assessment applied in cycle I.

B. The Washback Effects of Performance-Based Assessment on Student Learning

During the first cycle of activities, the students did the presentations in groups for 15 to 20 minutes. Before doing so, each group of students made some preparations by doing an intensive discussion with their peers in group. They discussed everything related to the materials that they wanted to present in front of the class. They had to prepare several slides of power point as well. During this preparation phase, it was observed that the students have shown a very good cooperative attitude because every group member could provide and receive their peers' opinions and thoughts. This cooperative attitude might certainly be one of the triggers to increase the student achievement. This is in line with the opinion of Gillies & Boyle (2010) who states that the cooperative attitude which emerged during the learning process can improve student academic achievement. Thus, cooperation among the students is very important in order to support the success of student learning.

In addition to the cooperative attitude, the discussions done by the students have also allowed the students to respect to each other. Each member of the group respects the thoughts and opinions put forward by others and seeks to respect the varieties of opinions and thoughts delivered by other group members as well. This mutual respect allowed the students to create a positive relationship among the group members so that the learning environment became more

conducive. Similarly, Bucholz & Sheffler (2009) states that the mutual respect among groups of students create a positive relationship among those students so that the learning environment will be more conducive. The cooperative attitude and mutual respect among the students were the positive washback effects arising from the implementation of performance-based assessment. The emergence of those students' behavior has directed the student learning became more productive and enjoyable.

After the preparatory process was finished, each group of students presented the group's work in front of the class. At the end of the presentation, there was a question and answer activity between the presenters and their audience and the questions should be in relation to everything that has been presented beforehand. The processes of question and answer were done spontaneously so that the students who asked the questions and the members of the group who answered those questions could communicate orally without using a variety of expressions and statements that had been prepared in advance. In other word, those students did oral communication in English spontaneously. The process of spontaneous oral communication was certainly capable of directing the students to be able to communicate better in English. Therefore, Buo and Waibel (2000) state that the process of spontaneous oral communication can train students to communicate with the target language automatically and independently. This statement implies that the process of spontaneous communication is capable of directing students to be able to communicate fluently in the target language.

It was, however, identified that the students' motivation to communicate orally in English observed in cycle I has not increased significantly. The results of the observations indicated that there were only a few students dared to express their thoughts, opinions, and ideas in the form of questions addressed to the group who did the presentation. Motivation was an important factor that might affect the success of student learning. Therefore, the students' motivation needed to be improved in cycle II so that their ability to speak in English might be improved as well. An effort to improve the students' motivation was then done by providing positive a stimulus in the form of additional scores for the students who dared to provide their thoughts, ideas and opinions. The stimulus was expected to increase the students' motivation to be better in the subsequent cycle.

During the second cycle, it was identified the students' motivation was much better than that of the students' motivation observed in cycle I. In one hand, the student's motivation identified in cycle I was still in a category of fairly high as only a few students who actively conveyed their ideas, opinions, and thoughts addressed to the members of groups who did a presentation. Most students still tended to be passive. In contrast, the students' motivation in the second cycle turned into high motivation in which most of students seizing an opportunity to convey their ideas, thoughts, and opinions in the form of questions addressed to members of other groups who had carried out the presentation. This suggested that the students' motivation in the second cycle turned into greatly higher. It occurred because there was positive stimulus given by the teacher in the form of the additional scores for each student who dared to express their ideas, thoughts and opinions orally addressed to the members of other groups who had done a presentation. Thus, it was understood that the stimulus that was rightfully provided to the students would be able to increase their motivation to learn more and more. Furthermore, Cheung (2001) states that the stimulus is a very effective key to the success of students' learning outcomes. Cheung also explains that the stimulus can increase students' motivation to learn English. Therefore, teachers need to provide a good stimulus in order to improve students' motivation to learn better and to increase the effectiveness of the process of learning as well.

Those analyses revealed several positive washback effects of performance-based assessment on secondary school student learning. Based on the results of in depth interviews and systematic observations, it was identified that performance-based assessment promoted positive washback effects on students' behavior and way of their learning with respect to the following matters (1) it eliminated the students' boredom and tedium during the speaking class; (2) it promoted the students' cooperative attitude; (3) it allowed the students for high order thinking as they had to spontaneously in English; (4) it increased the students' motivation to learn more and more; (5) it increased the students' quality of learning; (6) it provided the students a real-life communication in English; and (7) it caused to emerge the students' curiosity about the course contents. Therefore, those findings imply that oral presentation as one of performance-based assessments could be an effective pedagogical measurement tool which could well replace the traditional standardized achievement tests as these efforts were planned to maximize students' learning outcomes and students' development as well.

IV. CONCLUSION

This study revealed that performance-based assessment has positive washback effects on student learning in ELT classroom. These positive washback effects might not come about the traditional standardized achievement tests which are used and implemented in ELT classroom as the students just choose the options provided and are available in a particular set of the test. In other word, there might be significant difference between performance-based assessment and traditional standardized tests concerning their washback effect on the follow up learning of ELT classroom and performance-based assessment was proved to be of the positive washback effects while the traditional assessment modes were not comparatively of the same positive washback effect. Furthermore, the positive washback effect of performance-based assessment underscores the effectiveness of this language assessment mode as an alternative substitute of traditional assessment modes in educational measurement.

There are several implications for the results of the study. Firstly, English language teachers may need to reflectively think on their assessment practices to determine whether their assessment practices help to improve their students' language learning processes. Secondly, if teachers rightly assume that the main purpose of all assessment practices is to foster student earning, and if they further believe in the efficiency of the new alternative modes of assessment for the intended learning, it is justifiable to replace the traditional assessment procedures with the new alternative modes like performance-based assessment procedures.

REFERENCES

- [1] Abedi, J. (2010). *Performance assessments for English language teachers*. Stanford, CA: Stanford university, Stanford Center for Opportunity Policy in Education.
- [2] Aidogmus, Y; Aksu, R; & Kaya, M.F. (2014). The Relationship between Communication Skills of Geography Teachers and Students' Level of Motivation towards Geography Lesson. *International Journal of Academic Research in Business and Social Sciences* May 2014, Vol. 4, No. 5
- [3] Anderson, K. T. (2009). Applying positioning theory to the analysis of classroom interactions: Mediating micro-identities, macrokinds, and ideologies of knowing. *Linguistics and Education*, 20, 291–310.
- [4] Bass, K.M., Magone, M.E., & Glaser, R. (2002). *Informing the design of performance assessments using a content-process analysis of two NAEP science tasks*. (CSE Technical Report No. 564). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- [5] Bucholz, Jessica L. & Sheffler, Julie L. (2009). Creating a Warm and Inclusive Classroom Environment: Planning for All Children to Feel Welcome. *Electronic Journal for Inclusive Education*, Vol. 2, No. 4 [2009], Art. 4
- [6] Cheng, L., & Watanabe, Y. (2004). *Washback in Language Testing: Research context and methods*. New Jersey: Lawrence Erlbaum Associates, Inc.
- [7] Cheung, Chi-Kim. (2001). The use of popular culture as a stimulus to motivate secondary students' English learning in Hong Kong. *ELT Journal Volume 55/1 January 2001*
- [8] Darling-Hammond, L. (2006). No Child Left Behind and high school reform. *Harvard Educational Review*, 76(4), 642–667.
- [9] Doroban, D. (2007). *The methodology of evaluation and testing*. Athena: Ministerul Educatiei si Cercetarii
- [10] Gillies, Robyn M. & Boyle, Michael. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education* 26 (2010) 933-940.
- [11] Goldschmidt, P., Martinez, J. F., Niemi, D., & Baker, E. L. (2007). Relationships among measures as empirical evidence of validity: Incorporating multiple indicators of achievement and school context. *Educational Assessment*, 12(3 & 4), 239–266.
- [12] Linn, R.L., & Burton, E. (1994). Performance-based assessment: Implication of task specificity. *Educational Measurement: Issues and Practice*, 13(1), 5-8.
- [13] McTighe, J., & Ferrara, S. (2011). *Performance-based assessment in the classroom*. Retrived from <http://jwmcTighe.wordpress/wp-content/uploads/2011/004/performancebased-assessment-in-the-classroom.pdf>
- [14] Mueller, J. (2014). *Authentic assessment toolbox*. Retrieved from <http://jfmuellder.faculty.noctrl.edu/toolbox/whatisit.htm>
- [15] Pierce, L. V., & O'Malley, J. M. (1992). Performance and portfolio assessment for language minority students. *NCBE Program Information Guide Series*, Number 9, Spring 1992.
- [16] Quadir, M. (2014). Relationship between leaners' motivation factors and speaking strategy factors to learn oral communication in English. *The English Teacher* Vol XLIII (3) December 2014

Developing an Authentic Assessment Science Process Skills, Creative Thinking Skills and Manipulative Skills

Supahar¹, Dadan Rosana², Zamzam F A³, Ryani Andryani⁴, Neviana Wijayanti⁵

¹Faculty of Mathematic and Natural Science, Yogyakarta State University,

^{2,3,4,5}Science Education Study Program of Graduate School, Yogyakarta State University
supahar@uny.ac.id

Abstract - This research is aimed to know the procedure of instrument development of authentic assessment and to know the worthiness of authentic assessment instrument of development result is seen from the content validity by the validator. This research is a development research by model of non-test instrument. Development model of non-test instrument used has steps as follows: (1) determining of the instrument specification, (2) writing the instrument, (3) determining the instrument scale, (4) determining the scoring system, and (5) beating out the instrument. The writer used quantitative and qualitative technique to analyze the data obtained. The qualitative approach was used to analyze the input from experts and teachers, and the quantitative approach was used to analyze the results of experts' validation using Aiken's validity. Conclusion of this study are as follows: (1) The procedure of the development of authentic assessment follows the stages of research and development. The stages include pre-survey research, problem analysis, analysis of curriculum, research studies, experts' consultation, and drafting an instrument. The stages of development include experts' validation. (2) The quality of the develop products the developed authentic assessment has a valid criterion as an instrument, in terms of aspects of the construct, substance, and language. All these aspects meet a very good criterion and can be used with revisions.

Keywords: *authentic assessment, science process skills, creative thinking skills, manipulative skills*

I. INTRODUCTION

Natural Sciences is the mastery of facts, concepts, principles, and a process of discovery. The process of discovery in learning the natural sciences in accordance with the Nature of Science (NOS) means that science is a way of knowing. Lederman, et al. (1992: 231), stating that "that science is a way of knowing and there are values and beliefs inherent to the development of scientific knowledge". Based on these statements, NOS is defined as the concept of complex natural sciences involves philosophy, sociology, and historical knowledge.

Natural Sciences are the mastery of facts, concepts, principles, and a process of discovery. Learning the natural sciences is based on the contents of the standard form students who have a body of knowledge; standard process will shape the students with scientific skills, thinking skills and strategy of thinking; the standard scientific inquiry will form students capable of critical and creative thinking; as well as a standard assessment evaluates students humanely.

The learning process is directed at the development of the third realm of knowledge, attitudes, and skills should be implemented as a whole or holistically, meaning the development of one domain cannot be separated from other domains. The question that still occurs in the process of learning one's current assessment of the natural sciences still dominated the test form, which can only measure the realm of knowledge. The fact that learning the natural sciences is not always judged by using an assessment form test to measure student learning objectives. Assessment can be done by collecting information about students to give more accurate information about the skills and attitudes of students. The assessment directive can also be done to measure the learning process of students (Phopam 2008:6). That kind of assessment called the authentic assessment.

Authentic assessment is an assessment of immediate or direct size so that the assessment will be more obvious when votes directly to do with the granting of a task or project (Mueller 2006:1). Authentic assessment can be used to measure performance, achievement, motivation, and attitude of students in relevant activities in learning. The results of the study are eligible to be used as a basis in determining the kind of authentic assessment is (Stiggins, 1994:67): students' ability against (1) the substance of knowledge; (2) knowledge in doing the reasoning and solving problems; (3) skills in the mastery of knowledge; (4) the making of a product; and (5) achievement attitude in applying knowledge. The basic types of assessment methods offered by Stiggins (1994:83) include: (a) selected response assessment; (b) assessment essay; (c) performance assessment; and (d) personal communication assessment.

Portfolio, containers of evidence are becoming valuable tools for teacher and student assessment for reflection and metacognition, and for building collegial relationships (Collins, 1992: 451). Science process skills are all necessary to acquire, develop, and apply the concepts, laws, and theories of natural science, both in the form of mental skills, physical skills as well as social skills. The project assessment is an activity of the task of judging students from the stages of planning, implementation, and reporting that can develop creative thinking skills. According to Wang (2011: 1) defines creative thinking as the ability to sense problems, make guesses, generate new ideas, and communication results. Performance assessment requires students demonstrated their skills when performing the experiments so that can develop the manipulative skills. Manipulative skills i.e. the skill of preparing teaching materials and tools, take precautions and treatment (Das, 2007).

The material has different characteristics of natural sciences so not all matter natural sciences can be taught with the same method. Thus, the assessment instrument used of course will also be different, because if the instruments used are the same for all natural sciences material then there will be some aspects which cannot be measured. The selection of basic competence (KD) should be conducted to determine the appropriate type of assessment. In the development of this research material of the selected class VII natural sciences KD 2.3 can be used kind of an assessment portfolio. The assessment of the project can be used on a KD 3.8 with learning that directs students to develop creative thinking skills through assessment project. In addition, KD 3.6 performance assessment can be used to measure manipulative skills students in doing the experiment.

Referring to the problems outlined, then researchers trying to develop authentic assessment instrument can measure a few skills students i.e. science process skills, creative thinking skills and manipulative skills on some of the KD in the natural sciences learning in junior high school.

II. RESEARCH METHOD

A. Type of Research

This research included in the classification of research development. The products developed in this research in the form of instrument performance assessment, portfolio and project. Research development uses a five-step development instrument non test. Procedure of development following the stages of the development of non instrument test. Stages of the development of authentic assessments include (1) determining of the instrument specification, conduct an analysis of the specification of the instrument being developed include the analysis of students, needs analysis, analysis of curriculum, selecting the shape and format of the instrument, determine the indicators, making the latticework of instruments; (2) writing the instrument, writing of authentic assessment was developed based on the lattice that have been created and then draw up the details of the statement; (3) determining the instrument scale, the scale of the instrument that was used in the development of this authentic assessment instrument in the form of scales with a scale of 1 to 4; (4) determining the scoring system, a system of scoring in this authentic assessment instrument refers to the scale of use that is the scale of 1 to 4 to the emergence of student activities provided by the observer; and (5) beating out the instrument, perform the validation material, expert assessment and teacher.

B. Population and Sample

Population development of this authentic assessment is grade VII of the entire SMP/MTs in DIY. As for the samples used to involve grade VII of 3 SMP/MTs that is in DIY i.e., SMPN 2 Playen, SMPN 1 Piyungan and SMPN 2 Girisubo

C. Research Instrument

Data collection instruments used in this study consist of guidelines for interviewing sheet, observation sheet, and sheet now.

D. Data Analysis Techniques

Analysis of the validation of the content of the descriptive and quantitative basis. Quantitative analysis using V'aiken analysis (Azwar, 2014: 113) by the following formula:

$$V = \frac{\sum s}{n \cdot c - 1}$$

Description:

$s = r - lo$

n = number of panels of assessors

lo = lowest validity assessment

c = highest validity assessment

r = the numbers given by an assessor

III. RESULT AND DISCUSSION

A. Procedure the development of Authentic Assessments

The products developed are authentic assessment instrument which covers the instrument performance, portfolio, and projects. The instrument used to measure the performance of science process skills learners in the material system for excretion. The portfolio of instruments used to measure critical thinking skills learners on the material pressure of the liquid. Project assessment instrument used to measure problem solving skills learners on optical materials on the human eye. The assessment instrument developed is in the form of sheets of observations accompanied by grating and rubric assessments. Authentic assessment instruments development procedure is as follows.

1. Preliminary Studies

Some of the things done on the preliminary study include: analysis of the problem, an analysis of the curriculum, and the analysis of the learners. Problem analysis was done based on interviews with a number of teachers of science in SMP N 2 Playen, SMP N 2 Girisubo, and SMP N 1 Piyungan. The issues that emerged from the interviews that is not yet the availability of valid assessment instruments to measure skills learners, so it is important to develop these instruments. Curriculum analysis conducted to determine the competence of the basic curriculum of 2013 which corresponds to the selected material. Learner analysis aims to find out the characteristics of the students i.e. students of class VII junior high school.

Table 1. Curriculum Analysis Result

Types of Skills	Core Competence	Basic Competencies
Science Process Skills	3. Memahami pengetahuan (faktual, konseptual, dan prosedural) berdasarkan rasa ingin tahunya tentang ilmu pengetahuan, teknologi, seni, budaya terkait fenomena dan kejadian tampak mata. 4. Mengolah, menyaji dan menalar dalam ranah konkret (menggunakan, mengurai, merangkai, memodifikasi dan membuat) dan ranah abstrak (menulis, membaca, mengarang) sesuai dengan yang dipelajari di sekolah dan sumber lain yang sama dalam sudut pandang/teori.	3.9 Memahami konsep suhu, pemuaian, kalor, perpindahan kalor, dan penerapannya dalam mekanisme menjaga kestabilan suhu tubuh pada manusia dan hewan dalam kehidupan sehari-hari. 4.10. Melakukan percobaan untuk menyelidiki pengaruh kalor terhadap perubahan suhu dan perubahan wujud zat. 4.11. Melakukan penyelidikan terhadap karakteristik perambatan kalor secara konduksi, konveksi, dan radiasi.
Creative Thinking Skills	3. Memahami pengetahuan (faktual, konseptual, dan prosedural) berdasarkan rasa ingin tahunya tentang ilmu pengetahuan, teknologi, seni, budaya terkait fenomena dan kejadian tampak mata. 4. Mengolah, menyaji, dan menalar dalam	3.8 Mendeskripsikan interaksi antar makhlukhidup dan lingkungannya 4.12 Menyajikan hasil observasi terhadap ineraksi mahluk hidup dengan lingkungan sekitarnya

Types of Skills	Core Competence	Basic Competencies
	ranah konkret (menggunakan, mengurai, merangkai, memodifikasi, dan membuat) dan ranah abstrak (menulis, membaca, menghitung, menggambar, dan mengarang) sesuai dengan yang dipelajari di sekolah dan sumber lain yang sama dalam sudut pandang/teori	
Manipulatives Skills	3. Memahami pengetahuan (faktual, konseptual, dan prosedural) berdasarkan rasa ingin tahunya tentang ilmu pengetahuan, teknologi, seni, budaya terkait fenomena dan kejadian tampak mata 4. Mencoba, mengolah, dan menyaji dalam ranah konkret (menggunakan, mengurai, merangkai, memodifikasi, dan membuat) dan ranah abstrak (menulis, membaca, menghitung, menggambar, dan mengarang) sesuai dengan yang dipelajari di sekolah dan sumber lain yang sama dalam sudut pandang/teori.	3.6 Mengenal konsep energi, berbagai sumber energi, energi dari makanan, transformasi energi, respirasi, sistem pencernaan makanan, dan fotosintesis 4.8 Melakukan pengamatan atau percobaan sederhana untuk menyelidiki proses fotosintesis pada tumbuhan hijau. 4.9 Melakukan percobaan untuk menyelidiki respirasi pada hewan.

2. Determine the instrument's specifications

Science process skills indicators used in the assessment instrument are the prediction, measurement, experimentation, concluding, and communication. Indicators of creative thinking skills used in the instrument are analyzing, losing an idea, synthesizing, evaluating, creating, and visualizing. Indicators of manipulative skills used in the instrument teaching tool is to use correctly, clean the tool properly and store teaching tool teaching correctly.

3. Writing instruments

Writing instruments are carried out taking into account the aspect of material, construction, and language. The assessment instruments developed contain: title, usage instructions, scoring guidelines, e.g. scoring, and the observation sheet.

4. Determine the scale of the instrument and scoring system

The instrument was developed using a scale of 1-4. Scoring is determined in accordance with the scale used. The highest score of each indicator is 4 and the lowest is 1.

5. Reviewing instruments

Authentic assessment instruments developed were investigated by seven rater i.e. two expert lecturers and of five practitioners (teachers).

B. Results of Authentic Assessment Validation

Validation of product based on the assessment of the substance, construction, and language. Subsequent validation results are analyzed with the V'aiken approach that aims to quantify the magnitude of the content validity coefficient (V). The magnitude of the numbers V obtained confirmed with numbers based on table V'aiken. The minimum figure should be reached based on table V Aiken (1985: 134) category 4 range and number of panel 7 are 0,86. The magnitude of V is obtained on the validation of the portfolio assessment sheet to measure process skills in science is about 0,86-1. The magnitude of V is obtained on the project assessment sheet validation to measure creative thinking skills are about 0,90-1. The magnitude of V is obtained in the performance assessment sheet to measure manipulative skills is of 0,89-1.

Based on the results of the analysis of the magnitude of the content validity of the assessment instrument's third showed that magnitude V instruments already exceed the minimum coefficient of V'aiken. Thus, the assessment instruments developed meets the validity of the content. In addition to knowing the validity of the instrument developed, validation is aiming to obtain advice which can be used as material for the repair of the instrument before conducted trials at the school.

C. Revision of the product

Assessment instruments are revised based on some suggestions by experts and practitioners. During limited trials and operational field test or measurement is in not discovering things that demanded he do revision, so that the revision could be made only when the process of examination of the instrument. In more detail, some revisions to the product can be outlined as follows.

1. Revisions to the usage instructions of the instrument so that more communicative and clear.
2. Revision of the observation sheet so that each observation sheets are given examples of scoring.
3. Revision of the rubrics so that homogeneous and focus on the systematic sequence.
4. Details of revision of the statement so that the statements communicated and homogeneous with other grains in one indicator.
5. Details of revision of the statement on the indicators devised the hypothesis so that made that clear parameters for measuring the skills of learners and presented grain statement about the interconnectedness between variables.
6. The revised grain statement on indicators composing the purpose of probation order made clear parameters for measuring the skills of learners and presented a statement stating the presence of grains of the verb.
7. Revision details of a statement on the observation sheet so that the language clarified.
8. Revised assertions so that the grains statement made clear parameters for measuring the skills of learners.
9. Revised assertions so that the grains statement clarified.
10. Revision of the format of the observation sheet so that there are six columns on a sheet for granting score the learners are assessed.
11. Clarify how to use observational science process skills sheet that is by adding the phrase "give a sign check (√) in the number of students in student performance met observation of grain" on a

IV. CONCLUSION AND SUGGESTION

Conclusion of this study are as follows: (1) The procedure of the authentic assessment development follows the stages of research and development. The stages include pre-survey research, problem analysis, analysis of curriculum, research studies, experts' consultation, and drafting an instrument. The stages of development include experts' validation. (2) The quality of the developed products the developed authentic assessment has a valid criterion as an instrument, in terms of aspects of the construct, substance, and language. All these aspects meet a very good criterion and can be used with revisions.

REFERENCES

- [1] Azwar, S. (2014). *Reliabilitas dan Validitas, Edisi 4*. Yogyakarta: Pustaka Pelajar.
- [2] Collins, A (1992). *Portfolio for Science Education: Issues in purpose, structure, and authenticity*. Science Education 76 (4) 451-463.
- [3] Das, R. C. (2007). *Science Teaching in Schools*. New Delhi: Sterling Publisher.
- [4] Lederman, Norman S, G. (1992). *Students' and Teachers' Conceptions of the Nature of Science: A Review of the Research*. Journal of Research in Science Teaching. VOL. 29, NO. 4, PP. 331-359 (1992)
- [5] Mueller, John. (2006). *Authentic Assessment Tool Box*. From <http://jfmuellder.faculty.noctrl.edu/toolbox/index.htm> at 9 March 2015.
- [6] Popham, W.J. (2008). *Transformative Assessment*. USA: ASCD
- [7] Stiggins, R. J., (1994). *Student-Centered Classroom Assessment*. New York: Merrill, animprint of Macmillan College Publishing Company.
- [8] Wang, Amber Y. (2011). *Contexts of Creative Thinking: A Comparison on Creative Performance of Student Teachers in Taiwan and the United States*. Journal of International and Cross-Cultural Studies, Vol 2, Issue 1, 2011, ISSN 1984-5786.

Using of Self Assessment to Determine Science Process Skill and Concept Attainment Through Inquiri Learning of 8th Grade Student on 21th Junior High School in Ambon

Wa Nurlina¹⁾, K. Esomar²⁾, I. H. Wenno³⁾

^{1,2,3} Physics Education Programme-Faculty of Teaching and Education, Pattimura University
³ wennoiz@yahoo.co.id

Abstract - Inquiry learning, is one of learning models that explore the student ability to search and find the solution regarding the learning problem they have found during learning process. Inquiry learning emphasize the students to seek and find their own concepts to learn. This research uses descriptive method with the design of one group pretest-posttest design. Sampling with random sampling techniques, in order to obtain class VIII1 as other research samples. Data collection is the test, the assessment sheet in the form of self-assessment of performance, student work sheet, the observation sheet and questionnaire responses of students. The results showed that students 'science process skills and mastery of concepts students' achievement during the qualification process is good, (average 85.3) based on self-assessment, (82.6 based on an assessment observer) and average achievement of mastery of concepts students are 83.6. Posttest results show that science process skills and mastery of concepts students after the implementation of performance assessment with self-assessment techniques in inquiry learning is good categorized (an average of 83.9 students' science process skills), (average achievement of mastery concepts of students)

Keywords: *inquiry learning, self assessment, science process skill, concepts attainment*

I. INTRODUCTION

Assessment is one of important part of the learning process as well as in study of Natural Sciences [1]. According to Harlen [2] assessment is one aspect that can be used to improve the quality of education in the globalization era. Learning activities can be run with the maximum and growing scientific attitude of students. One of evaluation techniques that can be used is the self-assessment technique. Self-assessment is an assessment technique by asking learners to express themselves advantages and disadvantages in the context of the achievement of competencies [3]. According to Oscarsson cit (Basnet [4] self-assessment has been proven to have the number of benefits including awareness of student ability, motivation with purpose oriented, the promotion of learning, knowledge of valuation techniques, the various roles of assessment, increased responsibility for own learning, improve critical thinking ability of student, and improve the behavior of student. Self assesment can be applied through the inquiry learning. This is because the inquiry learning more emphasis on the investigation conducted by the student to obtain a solution of a problem. During inquiry learning students will gain experience of learning itself, then seek and find answers to the problems they face on a topic. By placing students at the center of learning (student centered learning), then students will have the ability to create, improve reasoning power, and developing science process skills.

Science process skills is a model wich can allow students to study something or problem based on scientific approaches by applying the scientific method including formulate the problem, propose and test hypotheses, making observations, define variables, designing and assembling the instrument, perform experiments, collect, process and interpret data, draw conclusions and communicate the results both orally and in writing. By developing science process skills, the students are able to master the concepts they have taught, so it will help improve students' mastery of the concept of what is learned [5,6].

However, in the process of teaching and learning in schools, teachers are generally not made a comprehensive assessment of student activity. Science process skills performed by students not only assessed through mastery of concepts based on the cognitive, but must be thoroughly covering both aspects of the psychomotor and affective development of students during science process skills in the inquiry learning. As in the 21th state Junior High School in Ambon. In this school, the assessment conducted by teacher still was limited to collecting a number of figures in the form of a score of student learning outcomes for a subject matter for a specific time period.

Generally, only the aspects that include students 'cognitive aspects of the students' knowledge, skills and psychomotor aspects while students are not involved in the assessment. According to Sudjana [7] the Natural Science learning objectives is summarized into three aspects, namely learning objective mastery of science concepts, skills development process or performance of students, and the planting of a scientific attitude. From these reason above it is needed an assessment to inquiry learning activities can be thoroughly developed three goals or dimensions of science learning the knowledge, skills and attitudes of learners. One of the most relevant to meet these demands is an authentic assessment of performance-based assessment (on performance assessment).

II. METHOD

This research is a descriptive study. The study design used is one group pretest-posttest design. This research was conducted at the 21st State Junior High School in Ambon. The population of all students of 8 grade consists of three classes. Sampling was done by random sampling, in order to obtain a sample that is in grade VIII₁. Instruments used to collecting data is the instrument tests for the science process skills and mastery concepts of students before and after the implementation of performance assessment. Sheets performance appraisal form of self-assessment to determine the performance of students in the form of science process skills of students during the learning process, students' worksheet to determine mastery of concepts during the process, the observation sheet to collect data during the process and is used as a benchmark assessment by students and questionnaire used to evaluate the response students to use the techniques of self-assessment and questionnaire responses of students.

Implementation of this learning begins with the provision granting initial test aims to determine how the initial process skills and mastery of the initial concept owned by students. Then the application of performance assessment with self-assessment techniques on inquiry learning. After that, final test was done to find out how the science process skills and mastery of concepts of students after learning process. Students are successful if individual achievement scores students achieve minimum completeness criteria benchmark scores of 70 and classically succeed if $\geq 70\%$ of students achieving a minimum completeness criteria benchmark scores.

III. RESULT AND DISCUSSION

Prior to learning by applying the techniques of self-assessment, measurement of science process skills and mastery of concepts students was done. The measurement results showed that overall science process skills and mastery of concepts students are on the category failed because the initial test score they obtained is less than 70 (Table 1 and Table 2).

Table 1. Result of Initial Test of Science Process Skill of Student

Mastery Level	Frequency	Percentation (%)	Qualification
90 – 100	-		Very good
80 – 89	-		Good
70 – 79	-		Medium
< 70	22	100	Failed
$(\bar{X}) = 44,8$			Failed

Table 2. Result Of Initial Test On Physics Mastery Concept Of Student

Mastery level	Frequency	Percentation (%)	Qualification
90 – 100	-		Very good
80 – 89	-		Good
70 – 79	-		Medium
< 70	22	100	Failed
$(\bar{X}) = 29$			Failed

After the initial test, learning process was done by applying a self-assessment techniques to measure the science process skills and mastery concepts of students. The learning process was conducted in two meetings. In each meeting self-assessment techniques was conducted both by students and by the observer. Rate observer is used as a benchmark for assessment of the student. The difference between a student assessments (self assessment) and observer ratings is shown in Figure 1.

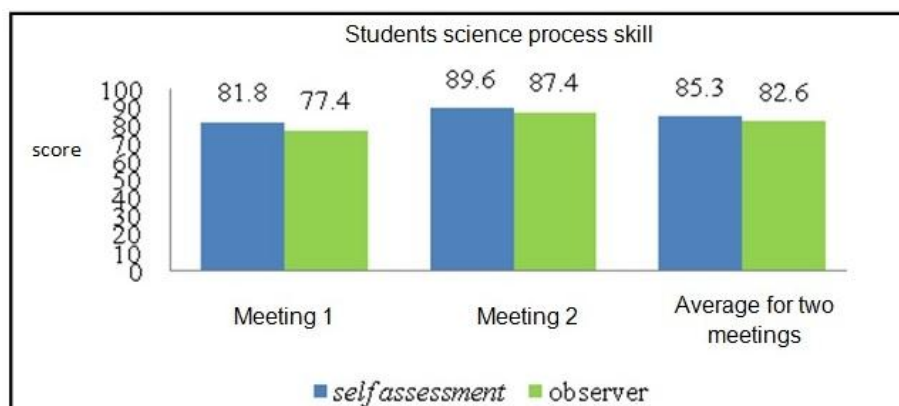


Figure 1. Student's science process skill after self assesment

Fig. 1 shows that the science process skills of students average is good categorized. But at the second meeting is higher than the first meeting. This means that the science process skills of students tends to increase with the implementation of self-assessment. According to Bound and Falchikov [8] stated that by using self-assessment techniques, students will evaluate the quality of their work and learning so this can allow the students to know the extent to which the goal is achieved, identify strengths and weaknesses in their work and improve student performance. Each stage of measurement of students science process skills were shown in Table 3.

Table 3. Each Step of Science Process Skill

Steps for Science Process skill	Percentation (%) meeting 1		Percentation (%) Meeting 2		Average percentation (%)	
	Self Assessment	Observer	Self Assessment	Observer	Self Assessment	Observer
Formulate problems	88	98	100	100	94	99
Formulate Hypotheses	83	80	92	92	88	86
Ability in using equipment and any other materials	92	79	95	91	94	85
Planning an experiment	83	77	88	86	86	82
Prediction	86	70	86	84	86	77
Make conclusion	70	73	91	88	80	80
Communication	70	65	70	70	70	67

Table 3 shows that the science process skills of students at every meeting was good categorized. This shows that the implementation of performance assessment with self-assessment techniques in inquiry learning was well done. However, for the ability to communicate is still considered enough, and this means the implementation of performance assessment with self-assessment techniques in inquiry learning has not impacted on the ability to communicate. The measurement results mastery of concepts during the process or during the implementation of self-assessment were shown in Figure 2.

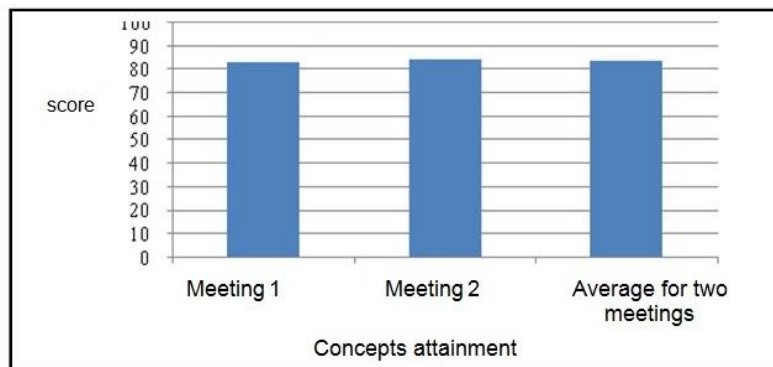


Figure 2. Students concepts attainment after self assesment

Figure 2 shows that the mastery of concepts for the implementation of performance assessment with self-assessment techniques are categorized either in inquiry learning. Overall, a good measurement science process skills Nor mastery of concepts for the implementation of self-assessment techniques shows the results of students are on good category. These results suggest that self assessment successfully improve science process skills and mastery concepts of students.

According to Noonan and Duncan [3] the self assessment techniques help students to define and determine the actions to be carried out students to develop their skills, and find the concept. Inquiry learning engages students in asking questions, searching for an explanation, the explanation and generate knowledge. By developing their science process skills, will help students to mastery the concepts they have learned [9]. Science process skills training can also improve motivation and student learning outcomes.

Orsmond [10] explained that by applying the self-assessment in the learning process, it will develop metacognitive ability of students. Further explained that when students develop their capacity to understand their own thought processes, they are more ready to use cognitive skills required to complete a task or achieve a goal.

IV. CONCLUSION

From the results of this study concluded that the implementation of performance assessment with self-assessment techniques in inquiry learning can help students recognize the science process skill and the achievement of students' mastery of concepts expressed both, and students gave a positive response to the use of self-assessment techniques.

REFERENCES

- [1] P. Race, "A briefing on self, peer and group assessment", 2001.
- [2] W. Harlen, "Assessment & Inquiry-Based Science Education: Issues in Policy and Practice". Global Network of Science Academies (IAP) Science Education Programme (SEP) 2013. www.interacademies.net/activities/projects/12250.aspx TWAS-Strada Costiera, 11-34151, Trieste, Italy
- [3] B. Noonan, and R. Duncan, "Peer and self-assesment in high school". Practical Assesment, Research and Evaluation journal. 2005, 10(17), pp. 1-8.
- [4] B. Basnet, M. Basson, C. Hobohm, and S. Cochrane. Students' Self-assessment of Assignments - is it Worth it?. In: Proceedings of the 2012 Australasian Association for Engineering Education (AAEE) Conference, Melbourne, Australia.
- [5] S. Ozgelen, "Students' Science Process Skills within a cognitive domain framework". Eurasia Journal of Mathematics, Science and Technology Education. 8(4), pp 283-292. 2012.
- [6] S. Karamustafaoglu, "Improving the science process skill ability of science student teachers using I diagrams" Eurasian Journal of Physics, Chemistry education, 3(1), pp 26-38. 2011.
- [7] N. Sudjana, "Penilaian Hasil Proses Belajar Mengajar" Bandung: Remaja Rosdakarya. 2011
- [8] D. Bound, N. Falchikov, Quantitative studies of student self-assessment in higher education: a critical analysis of findings. 1989, Volume 18, Issue 5, pp 529-549
- [9] M.L. Ango, Mastery of Science Process Skills and Their Effect Use in the Teaching of Science: An Educology of Science Education in the Nigerian Context. International Journal of Educology 16(1), pp 11-30. 2002.
- [10] P. Orsmond, Self and Peer-Assessment: Guidance on Practice in the Biosciences. 2011. Online available: <http://www.bioscience.heacademy.ac.uk> [30 Desember 2014].

DEVELOPMENT EVALUATION MODEL AND TECHNICAL EVALUATION MANAGEMENT PROGRAM MAHAD ALY IN THE COLLEGE OF ISLAMIC RELIGIOUS AFFAIRS (PTKIN)¹.

Winarno

State Institute for Islamic Studies (IAIN) Salatiga
winarno5@yahoo.co.id

Abstract - This research aims at: (1) producing evaluation model program Mahad Aly at College of Islamic Religious Affairs (PTKIN), and (2) producing technical evaluation management program Mahad Aly in the College of Islamic Religious Affairs (PTKIN). This research use the research and development approach (R&D) by using the nine steps of the 10-step of Borg and Gall model. Research sites in Mahad Aly the State Islamic University (UIN) Malang, and Mahad Aly State Islamic Institute (IAIN) Salatiga. Subject test consists of; manager, elements of leadership, and mahasantri. The data collections techniques were Delphi, focus group discussions, observation, interviews, documentations, and questionnaires. The collected data were analyzed through three types of activities going on simultaneously: data reduction, data presentation, and conclusion / verification. Qualitative analysis of the data using the display in the form of narrative, tables, and graphs. The findings of the research show: (1) evaluation model program ma'had Aly in order to improve the quality of program management at Mahad Aly College of Islamic Religious Affairs (PTKIN), among others; Improved means of learning, improvement of facilities, equipment, performance improvement and mentoring. (2) technical evaluation management program implementation of appropriate evaluation Mahad Aly management program at College of Islamic Religious Affairs (PTKIN) is; Rote maintenance techniques, techniques sustainability of the alumni participants ma'had, Regeneration - Regeneration, scientists scholars and clerics were scientists

Keywords: *evaluation model program, technical evaluation management program, and Mahad Aly*

I. A. INTRODUCTION

In the evaluation, if a program has been implemented it is needed to evaluate the program. Evaluation program is the process of systematically determination of the value, purpose, effectiveness or suitability something accordance with the criteria and goals. The decision process is based on a comparison carefully against the observed data by using specific standards that have been standardized.¹

Objective of evaluation program is "to revisit the achievement of objectives and to help provide the next alternative in the decision. With an evaluation then identified all the obstacles, the results of the evaluation are some tool recommendations for improvement, after the repair of the various sectors of the barrier have been completed, if the barriers have been completed. Diniyah education at higher education level can organize academic programs, vocational, and professional form of a university, college, or high school. The explanation of Article 20, paragraph 1 is education diniyah higher education among others Ma'had Aly. Ma'had Aly become one of the important phenomenon because Ma'had Aly in college combines Islamic studies in schools that specifically examine the treasures of classical Islamic scholarship with contemporary enriched material.

¹ This Research Was Funded From Quality Improvement Research Program the Directorate Of Higher Education Islamic Religious Ministry Indonesian Number of SK: 4692/2015 Date; August 18, 2015

Ma'had Aly formed in order to prepare a cadre of scholars who have scientific integrity, amaliah and khuluqiyah quality and strategic value-oriented justice, equality, openness, honesty, trustworthiness, and citizenship. Ma'had Aly based Ahlus Sunna Wal Jamaa'ah on the basis of Islam meant that Ma'had Aly held, organized, and developed a set off (point of departure) from Islamic teachings, Islamically management process and towards what is idealized by the Islamic education

A. Evaluation Program

Evaluation is the measurement, assessment and evaluation are hierarchical. The comparison of observation with the criterion is a measurement the interpretation and description of the evidence is an assessment and the judgment of the value of implication of the behavior is an evaluation. Meanwhile, according to the Joint Committee on Evaluation Standards describe that evaluation is the systematic assessment of the worth or merit of some object.ⁱⁱ

According Kufman and Thomas said that evaluation is a process used to assess. Another opinion defines the evaluation can be defined as the process of assessing something based on criteria or objective standards that are evaluated.ⁱⁱⁱ Evaluation is one of a series of activities to improve the quality, performance or productivity of an institution in implementing all the program. The purpose of evaluation is to see and know the processes that occur in the learning process. Through the evaluation will be obtained information about what has been achieved and what has not.^{iv} Evaluations provide information for classes and teachers to improve the quality of teaching and learning process. Evaluation as a component of teaching is a process to determine the success of teaching programs and an assessment process that aims to identify the difficulties inherent in the process of learning.^v

Evaluation of the definition above can be concluded that the evaluation is the systematic application of scientific procedures to assess the design, then presents the information in the decision making process of the implementation and effectiveness of a program. Ralph Tyler, 1950, program evaluation is a process to determine whether the program objectives can already be realized. Another opinion stated evaluation program is systematically setting process about the value, purpose, effectiveness or suitability something accordance with the criteria and goals set earlier. The decision process is based on a comparison carefully the observed data using specific standards that have been standardized.^{vi}

From the various definitions mentioned above, core of that the definition of program evaluation is an activity to gather information about the workings of something the government program, which further information is used to determine the appropriate alternative or choice in taking a decision.

B. Ma'had Aly

Ma'had Aly word by etymologically means high schools or in other words the college level. Name of ma'had aly for residential buildings because the students want to give a different impression. According to Imam Suprayoga, "hostel" connotes only as a place for students to move the bed. Not also called the "boarding school (*pesantren*)". Although culturally, the term "boarding school" can refer to "boarding school". The term is more emphasized that "boarding school" is not merely a "boarding school", a classic book of the Koran as general. But more than that, of the collaboration between salafi systems with modern systems.^{vii}

Ma'had Aly emergence was motivated by the scarcity of formal education that specifically scored scholars in a society that is undergoing change, although many colleges Islam. As is known in line with changes of modernization, people's lives and the nation of Indonesia continue to change and impact on diversity patterns of a more rational and functional. As the implications of this, is the clerical authorities have to deal with the various demands of society in a pragmatic livelihood.^{viii}

II. METHODS

A. Type and research approaches

This research is a research approach Research and Development (R & D). Step-by-step procedure using nine-step development of the teen-step of Borg & Gall model as Figure 1 below.

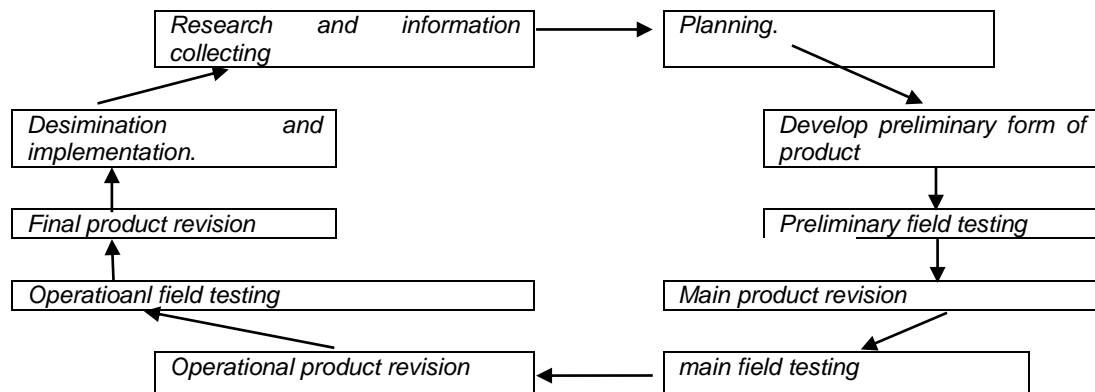


Figure 1. Step-step R & D research model of the development program evaluation and implementation techniques mahad aly in PTKIN

Based on Figure 1 above, research steps described as follows.

- Research and information gathering. The collection of information in this study is required to begin to gather detailed information and collect a number of problems found in the research in the field. The collection of information by conducting a literature review, reviewing some of the concepts of research and development (R & D), a preliminary investigation in two Ma'had Aly; are Ma'had Aly in UIN Maulana Malik Ibrahim Malang and Ma'had Aly IAIN Salatiga
- Planning includes planning products to be produced, the time of research, writing the initial draft;
- Development of the initial product form a form a model of evaluation, and technical implementation of the program, the subject of the trial as much as 4 test subjects consisting of two experts evaluation, two experts Ma'had management. Revision products, carried out repairs as has been proposed by a preliminary field test results;
- The field trials primary / expanded done in one Ma'had Aly is composed of two leaders of the boarding school, ma'had 4 lecturers and 10 students of the boarding school with the Delphi technique.
- Revision products based on feedback and suggestions as proposed by the main field test results;
- Operational field trials conducted in 2 Ma'had are Ma'had Aly in UIN Maulana Malik Ibrahim Malang and Ma'had Aly IAIN Salatiga. Subject test consists of leaders and managers ma'had of 4 people, 6 Ma'had Aly lecturers, and students 20 people.
- Revision of the final product, make improvements as proposed under the operational field test results so that the end product of the development (final product) in the form of a model development program evaluation and implementation techniques aly mahad management at College of Islamic Religious Affairs (PTKIN)

B. The data collection technique

Data collection techniques in this research through several techniques including:

- FGD techniques, this technique is a method of participation in the collection of information about a problem and needs very specific manner through discussion groups and to obtain mutual agreement. FGD can be used as a data collector or a research strategy;
- The first trial involving a small scale that represents the Ma'had Aly to test the device evaluation, evaluation model, and evaluation guidelines;
- Second trial on a larger scale involving one Ma'had Aly to test the device evaluation, evaluation model, and evaluation guidelines;

C. Methods of data collection research

Methods of data collection instruments using some sort;

- Observation to perform data to create a model development program evaluation and implementation techniques aly mahad management at College of Islamic Religious Affairs (PTKIN) and keep records in a systematic and logical, objective and rational;
- The interview is required to be able to reveal things that are still hidden, which was closed at the time to make observations about the model development program evaluation and

implementation techniques aliyah management at College of Islamic Religious Affairs (PTKIN)

- c. Documentation, this method is used to obtain information through document Ma'had Aliy. The data were analyzed by means mengorganisasikan data into categories, describe into the units, synthesize, organize into a pattern.

D. The analysis of experimental data

Analysis of the data used is descriptive qualitative data analysis is done by looking at the completeness of the model, evaluation model guide clarity, completeness and analysis of the effectiveness of the model instrument. The analysis model consists of four criteria that must be met, namely a comprehensive, practical, economical and supported by a valid and reliable instrument. Qualitative analysis of the data is the data menaganalisis validation results (assessment) of experts (expert) and the evaluation models (Ma'had Aliy) and practitioners who provide inputs for the improvement of evaluation models and devices. Qualitative analysis of the data using the display in the form of narrative, tables, and graphs covering the growing amount of data mahasiswa for three years, the output of students

III. DATA ANALYSIS AND DISCUSSION

The first, Model ma'had Aliy program evaluation in order to improve the quality of program management at Mahad Aliy College of Islamic Religious Affairs (PTKIN), among others; First, Repair learning tool for the learning process in Ma'had are not yet standardized. The amount of space for teaching and learning in Ma'had need to be improved and perfected for many desks are inadequate, lighting the lamp with less for the amount mahasiswa large and books Dirosah incomplete. Study table and standard lamps that will make residents Ma'had will be convenient to melaksanakan learning process. Besides, it is also less comprehensive learning tool with yet available places to study in a large room that can accommodate all residents Ma'had. Both Improvements in communications need to be improved. They often lack of water for bathing, washing, and toilet (MCK). Sports facility as a vehicle to make a fresh body because of the body that is fresh will make residents ma'had be able to think good and have strong endurance, it should be understood that the interaction with many occupants ma'had will cause occupants mah'had risky against certain diseases because all this is happening banya many schools by the number of students will be susceptible and infected with diseases such as colds, coughs, and skin diseases, third, performance improvement and assistance in implementing all the programs ma'had. Performance and passion for the success of the boarding school program needs to be improved with good discipline in every activity ma'had. Companion also needs to be maximized in the process of mentoring. In ma'had Arabic and English is still a problem if the percentage of partners and participants mah'had still not standard and well.

The Second, Technical implementation of appropriate evaluation Mahad Aliy management program at College of Islamic Religious Affairs (PTKIN) that needs to be done is; First, rote maintenance techniques so that participants can seal ma'had some juz Al-Qur'an as the target board and the boarding school program. Second, the survival techniques of alumni participants ma'had to simplify and expedite the boarding school program next. Ma'had program intended for new students and just one year or two semesters is still not adequate and long so it needs to be added for a few semesters. Third, Regeneration - Regeneration ma'had occupant is required to maintain continuity of the boarding school program. The cadres who need to be recruited to elect the next regeneration sehingga ma'had program could be continued and programs ma'had be more easily implemented. Fourth, the discourse of clerics and scholars scientist scientists as the main purpose of university education Islamic religious country should be supported by all parties in order to succeed. The dream that leaders in Indonesia country filled by alumni of PTKIN so should be encouraged to Sebera materialize, could have a president and vice president as well as ministers and other leaders who could memorize the Qur'an in addition to their degree.

IV. CONCLUSION

Model ma'had Aliy program evaluation in order to improve the quality of program management at Mahad Aliy College of Islamic Religious Affairs (PTKIN), among others; Improved means of learning, improvement of infrastructure, improvement of performance and mentoring.

Technical implementation of appropriate evaluation Mahad Aliy management program at College of Islamic Religious Affairs (PTKIN) is; Mechanical maintenance of foreign language skills, especially English and Arabic as well as memorization technique, technique sustainability

of the alumni participants ma'had, Regeneration - Regeneration, scientists scholars and clerics scientists

BIBLIOGRAPHY

-
- ⁱ Suharsimi Arikunto & Abdul Jabar. (2009). *Evaluasi Program Pendidikan: Pedoman Teoretis Praktis bagi Mahasiswa dan Praktisi Pendidikan*. Edisi Kedua. Jakarta: PT. Bumi Aksara, 14
- ⁱⁱ Stufflebeam, D.L. (1971). *Evaluation as enlightenment for decision making*. Columbus, Ohio: Ohio State University, 3
- ⁱⁱⁱ Djaali, Puji Mulyono dan Ramly. (2000). *Pengukuran Dalam Bidang Pendidikan*. Jakarta PPs UNJ, 3
- ^{iv} Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Press, 19
- ^v Murshel, J.L. (1954). *Successfull teaching, its psychological principles*. New York: Mc. Graw Hill Book Company Inc, 373
- ^{vi} Suharsimi Arikunto. (1988). *Penilaian program Pendidikan*. Jakarta: Bina Aksara, 25
- ^{vii} Taufiqurrochman, 2010, *Narasi Indah Perjalanan Hidup dan Pemikiran Prof. Dr. H. Imam Suprayogo*, Malang: UIN Maliki Press, 82
- ^{viii} Bagian Proyek Peningkatan Ma'had Aly, Naskah Kurikulum Ma'had Aly, Direktorat Pendidikan Keagamaan Dan Pondok Pesantren Direktorat Jenderal Kelembagaan Agama Islam Departemen Agama RI 2004, 4

THE DEVELOPMENT OF VOCATIONAL INTEREST INSTRUMENT FOR CAREER EXPLORATION OF JUNIOR HIGH SCHOOL STUDENTS

Yudhi Satria Restu Artosandi¹, Sudji Munadi²

¹ Faculty of Psychology, Universitas Muhammadiyah Surakarta

² Faculty of Technology, Universitas Negeri Yogyakarta

Abstract - The purpose of this study are: (1) to find out dimensions, aspects, and indicators of vocational interest as a basis for the establishment of vocational interest instrument measurement in the perspective of Holland Typology; (2) to develop model instrument of summated rating scale that can be applied to measure students vocational interest level; and (3) to establish the test of validity and reliability of vocational interest instrument and to test the measurement model for students of junior high school; (4) to produce a user manual and interpretation manual of vocational interest instrument for helping career exploration of junior high school students; (5) to produce a interest typology of the students who have studied in senior high schools and vocational schools. This study was a research and development (R & D) study. The research process was carried out through the stages of: (1) pre-development of theoretical review on the vocational interest concept and analysis of previous vocational interest relevant research; (2) development process consisting of a) arrangement of test specification and items of instrument, b) the evaluation by psychology and psychometric experts; c) analysis of tryout data quantitatively; and (3) instrument application, consisting of instrument application, scoring, and interpretation. The validity and reliability of the instrument were analyzed through *Confirmatory Factor Analysis (CFA)* in *Structural Equation Modeling (SEM)* using LISREL 8.80 Program. The results of the study indicate that: (1) the instrument is an inventory model of summated rating scale named as vocational interest scale containing 24 items describing vocational interest aspects; and (2) the instrument validity is significant as it shows that the lowest loading factor value is 0.61 and the highest is 0.89 (> cut-off value as 0.50). The reliability of the instrument is significant as the coefficient of reliability construct is 0.80 (>0.70). The model with six aspects of Holland typology including realistic, investigative, artistic, social, enterprising, and conventional (RIASEC) is fit because the *Mean Square Error of Approximation (RMSEA)* is 0.039 (< 0.08). The result shows that the model is suitable with the data. The model is able to estimate population covariance matrix which is not different from the sample covariance matrix so that the estimation result becomes a basis for generalization.

Keywords: vocational interest, instrument, measurement, validity and reliability

I. INTRODUCTION

The determination of next education level for junior high school graduates is important since it is presumed to be the early step toward their future career path. Yet it is somehow perplexing which compels them to seek for suggestion and recommendation from other parties, either from the school or other institutions. As a matter of fact, there is a limited psychological measurement tool that could be used to provide measurement to help the student optimally. There is a need for practical measurement that could serve as a tool for teachers to offer education consultant and initial counseling for their students. This measurement tool should be able to capture general information regarding with the suitability between student's characteristic and occupation preferences. One of prevailing measuring tools for career development is a theory from Holland which is developed based on Hexagonal theory that could help education practitioners in conducting counseling for career development. Based on Holland theory, an instrument for vocational interest can be established. It will explore the students' interest and

asses the career preference based on their experiences and interaction with their surroundings, thus, it will be beneficial for the students. The research aimed: 1) to establish an instrument of vocational interest based on Holland typology to support career exploration of the students in junior high school, 2) to establish the validation of vocational interest instrument based on Holland typology to support career exploration of the students in junior high school, 3) to provide guidelines for the use and interpretation of vocational interest instrument, 4) to provide interest typology for the students of senior high school and vocational high school which can be a standard reference of recommendation for higher education.

II. THEORETICAL REVIEW

A. Basic Assumption of Holland Personality Theory

The main focus of Holland theory is on the comprehension of vocational behavior to create a practical way to help the society, young, adult, and even senior citizens in starting their career, both in educational world or in working world (Louis, 2010). It emphasizes the concept of interest as the base of the formation of one's personality, as well as the personal competence, educational behavior, and social behavior and personality.

This theory is established based on four assumptions (Holland, 1997, p. 2-4) as follows: *First*, every person could be categorized based on to what extent they approach to one of the six personality types; *Second*, there are six environment models where each environment is dominated by a certain personality type and each environment has description of physical state, problems, and provide a certain opportunity and chance; *Third*, humans tend to look for an environment suitable to be used as the media in developing their expertise and capabilities, expressing attitude and value, and acquiring the correct and proper problem solution based on their characteristics. The combination of a certain personality and the suitable environment model will cause a harmony and occupational homogeneity, thus a person could perform self-development in a certain occupational environment and feels the satisfaction for it. *Fourth*, attitude is the embodiment of interaction between personality and environment. The suitability between individual and environment will determine the level of major conformity and stability of education and determining the satisfaction and achievement level.

B. Personality Types

Type is produced by type, meaning that even though parents' attitude have minimum and complex contribution for children's interest development (Roe, 1956; Roe and Siegelman, 1964 in Holland, 1997: 5), the assumption remains that parents' types will present activity environments towards their children based on the parents' types. Shortly, personality types according to Holland is the interaction result between hereditary factors and environment; and the interactions influence the preferences of certain types of activities, which will direct the individuals to certain attitude types. The summary of the six personalities is as follows:

1. Realistic type prefers on activities that will need explicit, orderly, and systemized manipulation towards objects, tools, machines, and animals. The implementation of this concept means that individuals with this type tend to dislike social or educational activities.
2. Investigative type prefers activities that will need observational, symbolic, systemized investigation and creative towards physical, biological, and cultural phenomena in order to understand and control the phenomena. It dislikes persuasive, social, and repetitive activities. The examples of occupations that will fulfill the needs of investigative type is chemists or physicists.
3. Artistic type prefers diverse, free, and unsystematic activities in order to produce artistic products such as painting, drama, prose. It dislikes systematic, orderly, and routine activities.
4. Social type prefers activities that will involve other people in the emphasis on helping, teaching, or providing social services. It dislikes routine and systemized activities involving objects and material objects.
5. Enterprising type prefers activities that involve manipulation of other people in order to acquire economic satisfaction and organizational goals. It dislikes systemized, abstract, and scientific activities.
6. Conventional type prefers activities with explicit, orderly, systemized data manipulation in order to provide contribution for organization's goals. It dislikes uncertain, free, and unsystematic activities.

III. RESEARCH METHODOLOGY

Operationally, this research follows the following steps:

1. Conducting introductory research, by doing literature and research review on the satisfaction level of students, parents, and teachers in the vocational process conducted by senior high schools.
2. Developing research design, complemented by frameworks on the initial step.
3. Developing research instrument.
4. Developing instrument of vocational interest used to determine the majoring process in senior high schools. The creation of interest measurement tools is predicted to be implemented in the determination of students' major.
 - a. Process and describe the findings of the introductory research. The data acquired of the introductory research are basic data of empirical studies, especially the ones related to the administration of majoring commonly done by senior high schools.
 - b. Studying the report of the administration of majoring process in some senior high schools, to be used as reference for the development of conceptual model.
 - c. Reviewing some theories and concepts to be used as references in the development of measurement instrument of vocational interest as the author's framework of thinking.
 - d. Creating drafts for measurement instrument of vocational interest based on empirical and conceptual studies.
 - e. Conducting limited discussion with some practitioners regarding the soon-to-be-developed measurement instrument of vocational interest.
 - f. Revising drafts for measurement instrument of vocational interest to the supervisory teacher and the expert of education.
5. Conducting validation for measurement instrument of vocational interest towards colleagues, supervisory teacher, and the experts in interest development field.
6. Revising measurement instrument of vocational interest based on the suggestions given by experts and program organizers of majoring process in senior high school.
7. Conducting a test for measurement instrument of vocational interest in the addressed field in order to create measurement instrument of vocational interest with high goodness of fit with the theories used.
8. Perfecting the instrument, through the data processing and analysis, and revising the formulas. The stage of refinement of data model is acquired through post-test result, field notes, discussion result, interview result, and documentation.
9. Compiling research report, as the final activity of research and development process.
10. Dissemination and distribution of instrument to be directly implemented to senior high schools.

This study was conducted in several senior high schools in Surakarta, Boyolali, Pacitan and Purwokerto with total of 900 respondents. It was a qualitative research using structural equation model (SEM). It is preferred since it is combination between confirmatory factor analysis and path analysis (Hadi, 2009). The calibration of structural equation model is by using Lisrel program with strictly confirmatory model to determine one constructed model and collecting empirical data to test the existing model. It is similar with a research conducted by Wong and Wong (2009) that implemented CFA to test the suitability of the model and the underlying theories. The content validity of Vocational Interest Indonesian Version (VIIV) is analyzed through exploratory factor on 60 items with eigen values = 1 and loading factor = 0.4. The calibration of model suitability towards RIASEC typology theory is by using Structural Equations Modeling approach to correlate with the main dimension. Comparison with RIASEC typology is conducted by examining the compatibility between the developed models in instrument VIIV with hexagonal model of structural hypothesis of Holland's theory, while the adjacent types have high correlation and contrasting types, indicated by far distance between both types in hexagonal model, have low correlation. In addition, the indication of goodness of fit used in the research to test the compatibility between VIIV with RIASEC model is by using RMSEA Steiger-Lind, Noncentrality index from McDonald, Population Gamma Index, Joreskog GFI and AGFI and Chi-squared goodness of fit test. If RMSEA criteria used is less than 0.05, then it indicates as highly fit; if it is between 0.05 – 0.08, it shows that it is fit; 0.08 – 0.10 shows medium fit; and if the value is bigger than 0.10, it shows misfit (Darcy & Tracey, 2007). GFI value is the indication of variances, which is measured from the model and GFI index is on between 0 (for badfit) to 1 (for perfect fit). The reliability indicator reflected from square multiple correlations (R²) with general requirement ≥ 0.40 , which shows variance proportion of each indicator that could be explained by its underlying factor.

IV. RESULTS

The data for experiment collection process was conducted towards senior high school students in some schools in Surakarta region. The schools were derived from random sampling selection of schools in Surakarta, Pacitan, Purwokerto and Boyolali, and there were 215 respondents. It was conducted on 1 May 2014. The validation of instrument was conducted by testing all dimensions of total six Holland typology dimensions in vocational interest measurement tool. The first step is testing *Realistic* dimension, which the findings could be seen in the following part.

A. CFA with Realistic dimension

Measurement model of *confirmatory factor analysis* with *realistic* (R) dimension is shown through the following *standardized* figure:

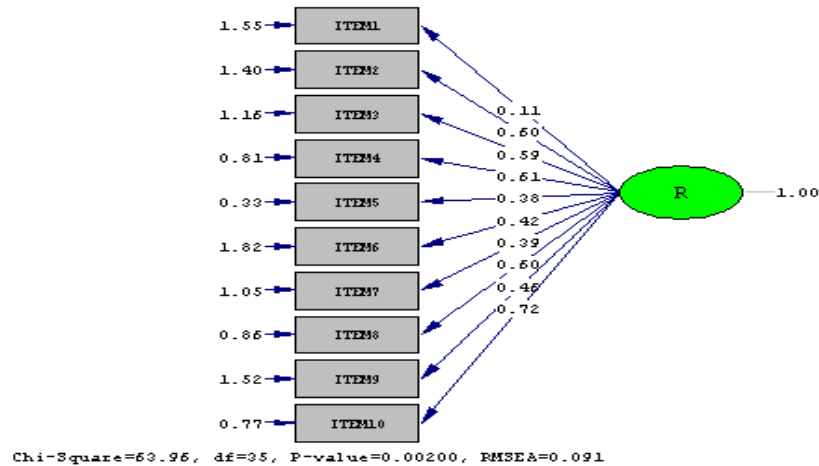


Figure 1. Result of *Standardized Solution* with Realistic Dimension

The model on Figure 5 shows the size of *loading factor* in each indicator towards its latent variable.

Table 1. Evaluation of Factor Loading and Criteria of Overall Measurement Model
Fit *Realistic* (R) Dimension

Indicator	Factor Loading	t – value (Critical Ratio)
Item1	0.11	0.74
Item2	0.60	4.12
Item3	0.59	4.39
Item4	0.61	5.28
Item5	0.38	5.15
Item6	0.42	2.66
Item7	0.39	3.18
Item8	0.60	5.05
Item9	0.46	3.15
Item10	0.72	6.08
Goodness of Fit Indices	Value	Decision
p value	0.002	Misfit
RMSEA	0.091	Misfit

The result of statistic measurement on figure 5 and table 3 shows that not all indicators of Realistic dimensional measurement have factor loading bigger than 0.5, which are item1, item5, item6, item7 and item9; however, most of them are significant ($t > 1.96$) except for item1 with t -value = 0.74. Observed from the result, the developed model is seen as misfit, which means zero hypothesis is rejected, which also means that the hypothesized model is dissimilar with the empirical data. Since the model is misfit ($p < 0.05$), it is important to do model modification. The suggested modification is by looking at the modification indices, which provide information on

the correlation between indicators of a latent construct. Based on the output, there is a correlation among the error variance on item4, item5, and item6. The correlation between error variance informs that the indicators are strongly inter-correlated and explains one similar value related to Realistic dimension. If they are observed, item4, item5 and item6 turn out having opposing sentence meaning, but the three of them are still the indicators of the similar dimension. If the three items are omitted from the model, the result could be seen in the following picture.

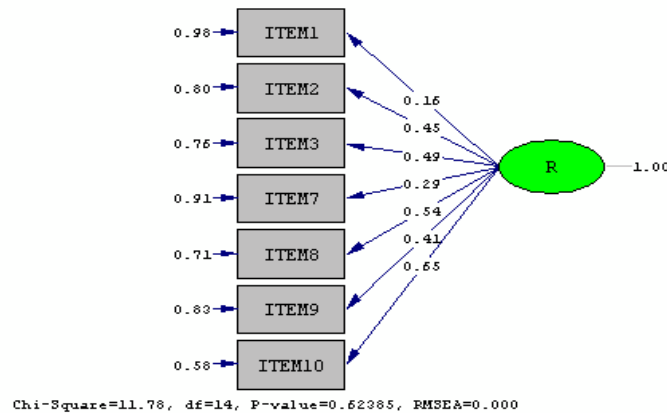


Figure 2. Result of *Standardized Solution* Realistic Dimensional Measurement Model (Revised Model)

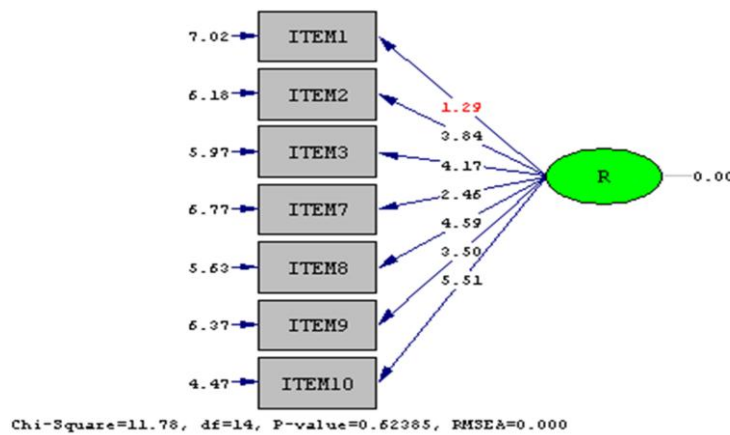


Figure 3. Result of *t-value* Realistic Dimensional Measurement Model (Revised Model)

The model (in the picture) shows the amount of *loading factor* of each indicator towards its respective latent variables. The amount of loading factor of each indicator is presented in the following table.

Table 2. Evaluation of Factor Loading and Criteria of Overall Measurement Model Fit Realistic (R) Dimension (Revised Model)

Indicator	Factor Loading	t – value (Critical Ratio)
Item1	0.16	1.29
Item2	0.45	3.84
Item3	0.49	4.17
Item7	0.29	2.46
Item8	0.54	4.59
Item9	0.41	3.50
Item10	0.65	5.51
Goodness of Fit Indices	Value	Decision
p value	0.62385	Fit
RMSEA	0.0000	Fit

Figure 6, figure 7, and table above present that by omitting three items, it could produce *good fits* since $p = 0.62385$ ($p > 0.05$) and $RMSEA = 0.0000$ ($RMSEA < 0.05$) (Darcey & Tracey in Louis, 2010). Generally, if the t value is observed carefully, all items will be bigger than 1.96 except for item 1, which means that the indicator will conform to Realistic theory concept in RIASEC model. Therefore, it is better to omit item 1 in the use of this dimension because the t value is < 1.96 .

B. CFA with Investigative dimension

Measurement model of *confirmatory factor analysis* with *Investigative(R)* dimension is shown through the following picture:

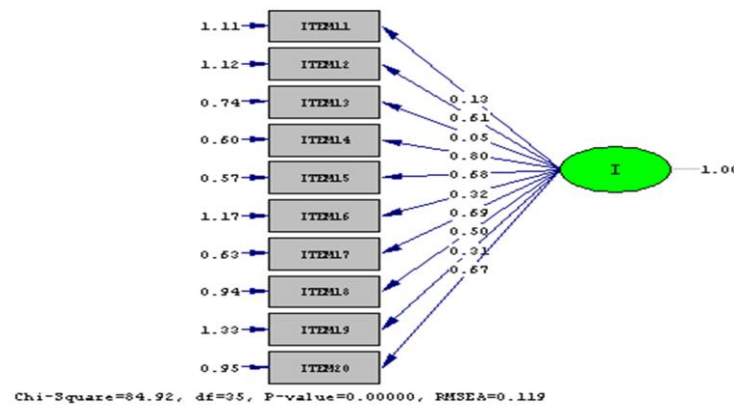


Figure 4. Result of *Standardized Solution* Investigative Dimension

The model on Figure 8 shows the size of *loading factor* in each indicator towards its latent variable. The size of loading factor of each indicator is presented in the data of table 3:

Table 3. Evaluation Factor Loading and Criteria of Overall Measurement Model Fit Investigatic (I) Dimension

Indicator	Factor Loading	t – value (Critical Ratio)
Item11	0.13	1.09
Item12	0.61	4.77
Item13	0.05	0.48
Item14	0.80	7.35
Item15	0.68	6.76
Item16	0.32	2.62
Item17	0.69	6.63
Item18	0.50	4.38
Item19	0.31	2.39
Item20	0.67	5.58
Goodness of Fit Indices	Value	Decision
p value	0.000	misfit
RMSEA	0.119	misfit

The result of statistic measurement on figure 8 and table 3 shows that not all indicators of Investigative dimensional measurement have factor loading bigger than 0.5, which are item11, item13, item16, and item19; however, most of them are significant ($t > 1.96$) except for item11 with t -value = 0.74 and item13 with t -value = 0.48. Observed from the result, the developed model is seen as misfit, which means zero hypothesis is rejected, which also means that the hypothesized model is dissimilar with the empirical data. Since the model is misfit ($p < 0.05$), it is important to do model modification. The suggested modification is by looking at the modification indices, which provide information on the correlation between indicators of a latent construct. Based on the output, there is a correlation among the error variance on item 14, item 15, item16, item18, and item 19. The correlation between error variance informs that the indicators are strongly inter-correlated and explains one similar value related to Investigative dimension. If they

are observed, item 14, item 15, item 16, item 18, and item 19 turn out having opposing sentence meaning, but all of them are still the indicators of the similar dimension. If the five items are omitted from the model, the result could be seen in the following picture.

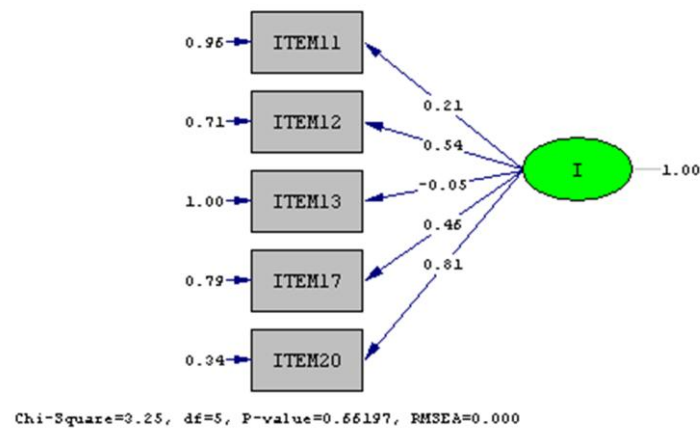


Figure 5. Result of *Standardized Solution* Investigative Dimensional Measurement Model (Revised Model)

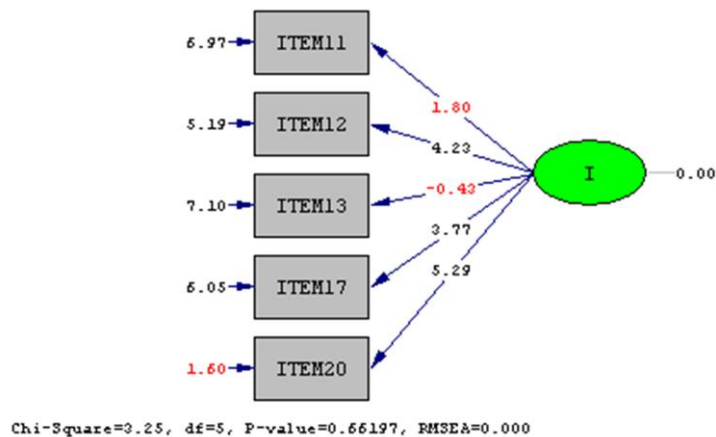


Figure 6. Result of *t-value* Investigative Dimensional Measurement Model (Revised Model)

The model (in the picture) shows the amount of *loading factor* of each indicator towards its respective latent variables. The amount of loading factor of each indicator is presented in table 4:

Table 4. Evaluation of *Factor Loading* and Criteria of *Overall Measurement Model Fit Investigative (I) Dimension* (Revised Model)

Indicator	Factor Loading	t – value (Critical Ratio)
Item11	0.21	1.80
Item12	0.54	4.23
Item13	-0.05	-0.43
Item17	0.46	3.77
Item20	0.81	5.29
Goodness of Fit Indices		Decision
p value	0.6619	<i>Fit</i>
RMSEA	0.0000	<i>Fit</i>

Figure 9, figure 10, and table 4 above present that by omitting five items, it could produce *good fits* since $p = 0.6619$ ($p > 0.05$) and $RMSEA = 0.0000$ ($RMSEA < 0.05$) (Darcey & Tracey in Louis, 2010). Generally, if it is observed carefully, all items will have t value bigger than 1.96 except for item13, which means that the indicator will conform to Investigative theory

concept in RIASEC model. Therefore, it is better to omit item 1 in the use of this dimension because the t value is < 1.96 .

C. CFA with Artistic dimension

Measurement model of *confirmatory factor analysis* with Artistic (A) dimension is shown through the following picture:

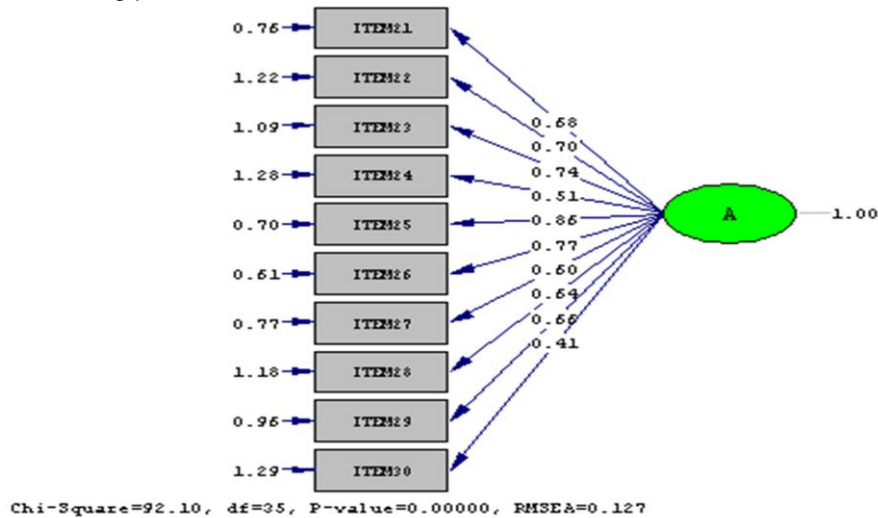


Figure 7. Result of Standardized Solution Artistic Dimension

The model on Figure 7 shows the size of *loading factor* in each indicator towards its latent variable. The size of loading factor of each indicator is presented in the data of table 5.

Table 5. Evaluation of Factor Loading and Criteria of Overall Measurement Model Fit Artistic (A) Dimension

Indicator	Factor Loading	t – value (Critical Ratio)
Item21	0.678	6.289
Item22	0.703	5.352
Item23	0.743	5.867
Item24	0.512	3.968
Item25	0.863	7.688
Item26	0.773	7.514
Item27	0.601	5.696
Item28	0.637	4.996
Item29	0.658	5.601
Item30	0.410	3.223
Goodness of Fit Indices	Value	Decision
p value	0.000	misfit
RMSEA	0.127	misfit

The result of statistic measurement on figure 11 and table 5 shows that all indicators of Artistic dimensional measurement have factor loading bigger than 0.5, except item30 which have factor loading = 0.410; however, all of them are significant ($t > 1.96$). Observed from the result, the developed model is seen as misfit, which means zero hypothesis is rejected, which also means that the hypothesized model is dissimilar with the empirical data. Since the model is misfit ($p < 0.05$), it is important to do model modification. The suggested modification is by looking at the modification indices, which provide information on the correlation between indicators of a latent construct. Based on the output, there is a correlation among the error variance on item21, item22, item 23, item28, and item29. The correlation between error variance informs that the indicators are strongly inter-correlated and explains one similar value related to Artistic dimension. If they are observed, item21, item22, item23, item28, and item29 turn out having opposing sentence meaning, but the five of them are still the indicators of the similar

dimension. If the five items are omitted from the model, the result could be seen in the following picture.

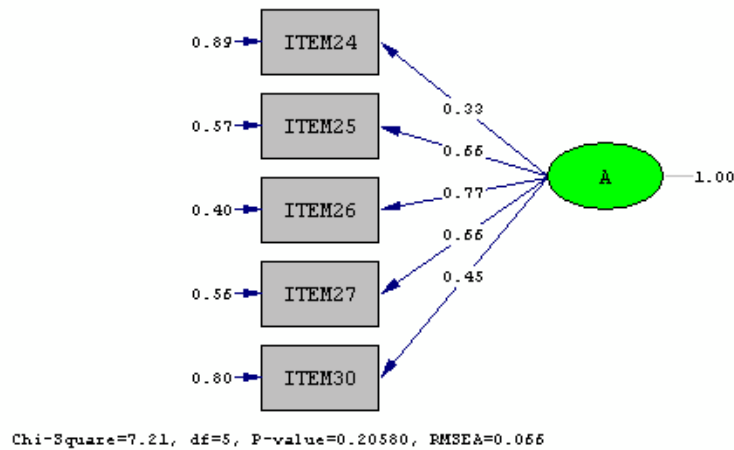


Figure 8. Result of *Standardized Solution* Artistic Dimensional Measurement Model(Revised Model)

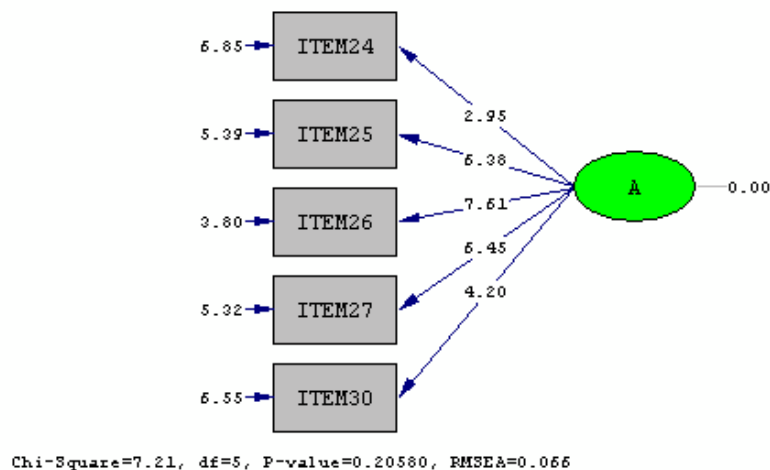


Figure 9. Result of *t-value* Artistic Dimensional Measurement Model(Revised Model)

The model (in the picture) shows the amount of *loading factor* of each indicator towards its respective latent variables. The amount of loading factor of each indicator is presented in table 6:

Table 6. Evaluation of *Factor Loading* and Criteria of Overall Measurement Model Fit Artistic (A) Dimension (Revised Model)

Indicator	Factor Loading	t – value (Critical Ratio)
Item24	0.33	2.95
Item25	0.66	6.38
Item26	0.77	7.61
Item27	0.66	6.45
Item30	0.45	4.20
Goodness of Fit Indices	Value	Decision
p value	0.2058	<i>Fit</i>
RMSEA	0.0066	<i>mediocre fit</i>

Figure 11, figure 12, and table 6 above show that by omitting five items, it could produce *good fit* since $p = 0.2058$ ($p > 0.05$) and $RMSEA = 0.0066$ ($RMSEA < 0.08$) (Darcey & Tracey in Louis, 2010). Generally, if the t value is observed carefully, all items will have t value bigger than 1.96, which means that the indicator will conform to Artistic theory concept in RIASEC model.

D. CFA with Social dimension

Measurement model of *confirmatory factor analysis* with *Social(S)* dimension is shown through the following picture:

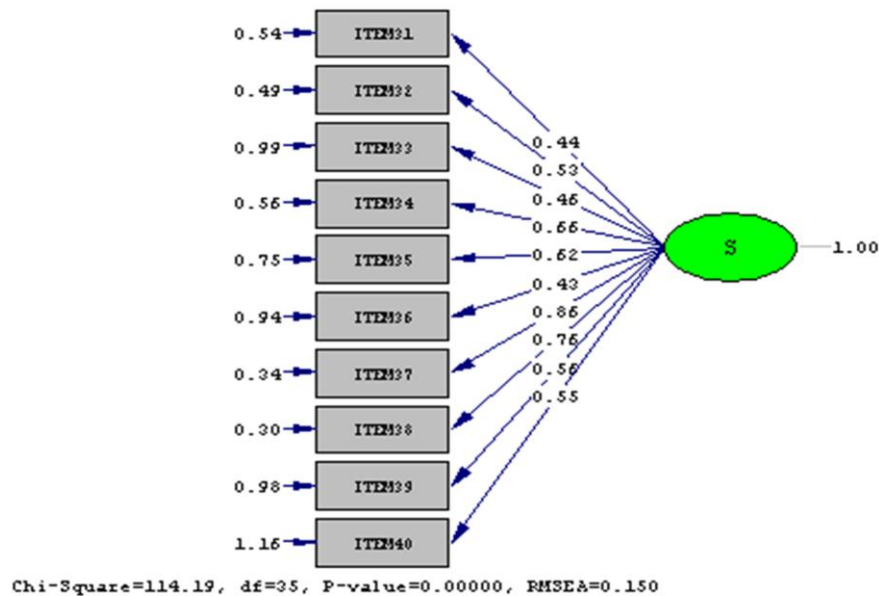


Figure 10. Estimation Result of Social Dimension

The model on Figure 10 shows the size of *loading factor* in each indicator towards its latent variable.

Table 7. Evaluation of *Factor Loading* and Criteria of *Overall Measurement Model Fit Social (S) Dimension*

Indicator	Factor Loading	t – value (Critical Ratio)
Item31	0.442	5.254
Item32	0.531	6.362
Item33	0.459	4.142
Item34	0.658	7.084
Item35	0.618	6.053
Item36	0.428	3.975
Item37	0.826	9.740
Item38	0.756	9.392
Item39	0.557	4.942
Item40	0.547	4.519
Goodness of Fit Indices	Value	Decision
p value	0.000	misfit
RMSEA	0.150	misfit

The result of statistic measurement on figure 13 and table 7 shows that all indicators of Social dimensional measurement have factor loading bigger than 0.5, except item31, item33, and item36 which have factor loading < 0.50; however, all of them are significant ($t > 1.96$). Observed from the result, the developed model is seen as misfit, which means zero hypothesis is rejected, which also means that the hypothesized model is dissimilar with the empirical data. Since the model is misfit ($p < 0.05$), it is important to do model modification. The suggested modification is by looking at the modification indices, which provide information on the correlation between indicators of a latent construct. Based on the output, there is a correlation among the error variance on item31, item 32, item 33, item 35, item 37, and item38. The correlation between error variance informs that the indicators are strongly inter-correlated and explains one similar value related to Social dimension. If they are

observed, item31, item32, item33, item35, item37, and item38 turn out having opposing sentence meaning, but the six of them are still the indicators of the similar dimension. If the six items are omitted from the model, the result could be seen in the following picture.

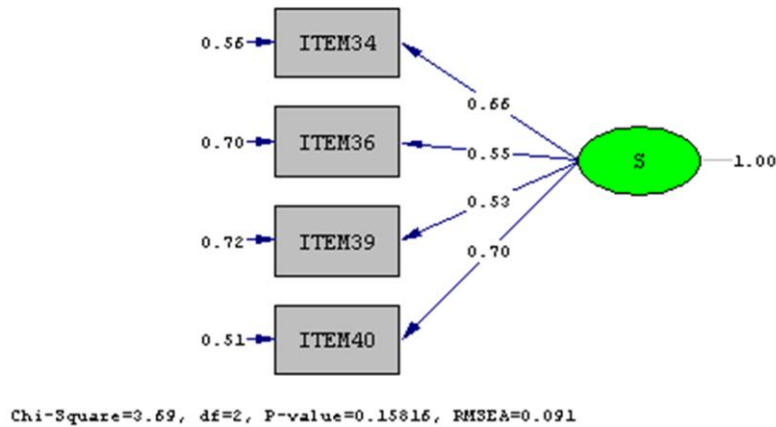


Figure 11. Result of *Standardized Solution* Social Dimensional Measurement Model (Revised Model)

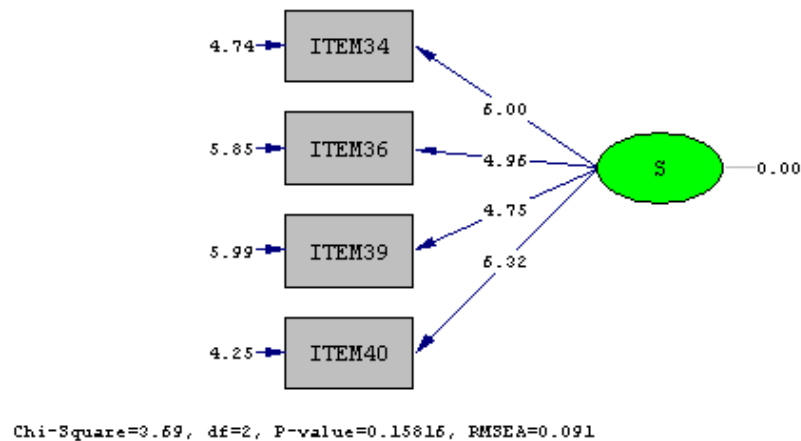


Figure 12. Result of *t-value* Social Dimensional Measurement Model (Revised Model)

The model (in the picture) shows the amount of *loading factor* of each indicator towards its respective latent variables. The amount of loading factor of each indicator is presented in the following table:

Table 8. Evaluation of Factor Loading and Criteria of Overall Measurement Model Fit Investigatic (I) Dimension (Revised Model)

Indicator	Factor Loading	t – value (Critical Ratio)
Item34	0.66	6.00
Item36	0.55	4.96
Item39	0.53	4.75
Item40	0.70	6.32
Goodness of Fit Indices		Decision
p value	0.15816	Fit
RMSEA	0.0091	Fit

Figure 14, figure 15, and table 8 above show that by omitting six items, it could produce *good fits* since $p = 0.15816$ ($p > 0.05$) and $RMSEA = 0.091$ ($RMSEA < 0.05$) (Darcey & Tracey in Louis, 2010). Generally, if the t value is observed carefully, all items will have t value bigger

than 1.96, which means that the indicator will conform to Social theory concept in RIASEC model.

E. CFA with Enterprising dimension

Measurement model of *confirmatory factor analysis* with *Enterprising*(E) dimension is shown through the following picture:

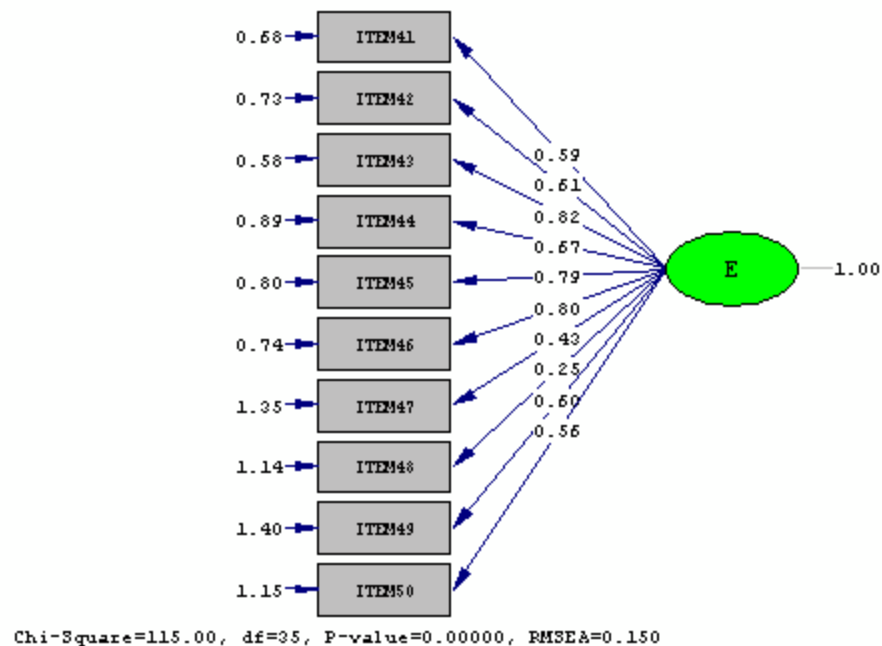


Figure 13. Estimation Result of Enterprising Dimensional Measurement Model

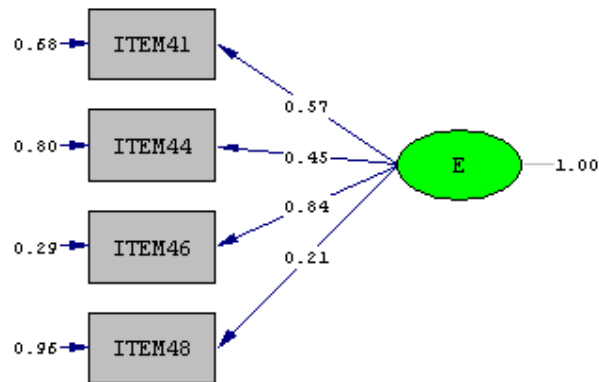
The model on Figure 13 shows the size of *loading factor* in each indicator towards its latent variable. The size of loading factor of each indicator is presented in the data of table 9:

Table 9. Evaluation of Factor Loading and Criteria of Overall Measurement Model Fit *Enterprising* (E) Dimension

Indicator	Factor Loading	t – value (Critical Ratio)
Item41	0.585	5.842
Item42	0.614	5.903
Item43	0.820	7.904
Item44	0.669	5.842
Item45	0.794	6.925
Item46	0.796	7.136
Item47	0.434	3.318
Item48	0.251	2.129
Item49	0.600	4.542
Item50	0.563	4.542
Goodness of Fit Indices	Value	Decision
p value	0.000	misfit
RMSEA	0.150	misfit

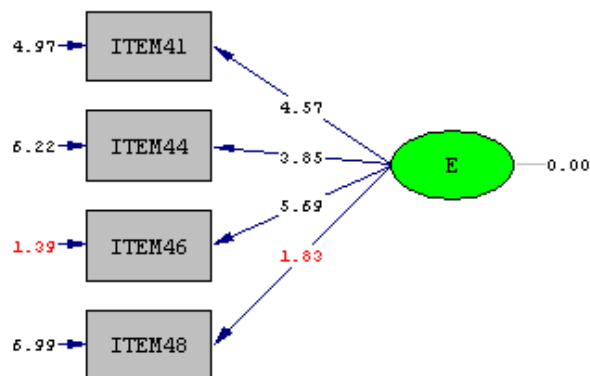
The result of statistic measurement on figure 16 and table 9 shows that all indicators of Enterprising dimensional measurement have factor loading bigger than 0.5, except item47 and item48 which have factor loading < 0.50; however, all indicators are significant ($t > 1.96$). Observed from the result, the developed model is seen as misfit, which means zero hypothesis is rejected, which also means that the hypothesized model is dissimilar with the empirical data. Since the model is misfit ($p < 0.05$), it is important to do model modification. The suggested modification is by looking at the modification indices, which provide information on the

correlation between indicators of a latent construct. Based on the output, there is a correlation among the error variance on item42, item43, item45, item47, item49, and item50. The correlation between error variance informs that the indicators are strongly inter-correlated and explains one similar value related to Enterprising dimension. If they are observed, item42, item43, item45, item47, item49, and item50 turn out having opposing sentence meaning, but the six of them are still the indicators of the similar dimension. If the six items are omitted from the model, the result could be seen in the following picture.



Chi-Square=4.74, df=2, P-value=0.09360, RMSEA=0.116

Figure 14. Result of *Standardized Solution* Enterprising Dimensional Measurement Model (Revised Model)



Chi-Square=4.74, df=2, P-value=0.09360, RMSEA=0.116

Figure 15. Result of *t-value* Enterprising Dimensional Measurement Model (Revised Model)

The model (in the picture) shows the amount of *loading factor* of each indicator towards its respective latent variables. The amount of loading factor of each indicator is presented in the following table:

Table 10. Evaluation of *Factor Loading* and *Criteria Overall Measurement Model Fit Enterprising (E) Dimension* (Revised Model)

Indicator	Factor Loading	t – value (Critical Ratio)
Item41	0.57	4.57
Item44	0.45	3.85
Item46	0.84	5.69
Item48	0.21	1.83
Goodness of Fit Indices	Value	Decision
p value	0.0936	<i>Fit</i>
RMSEA	0.116	<i>Misfit</i>

Figure 15, figure 16, and table 10 above present that by omitting six items, it could produce good fitsince $p = 0.0936$ ($p > 0.05$); however, it will be misfit on $RMSEA = 0.116$ ($RMSEA < 0.05$) (Darcey & Tracey in Louis, 2010). Generally, if the t value is observed carefully, all items will have t value bigger than 1.96 except for item48, which means that the indicator will conform to Enterprising theory concept in RIASEC model. Therefore, if the model is to be used, it is better to omit item48 off the model.

F. CFA with Conventional Dimension

Measurement model of *confirmatory factor analysis* with *Conventional(C)* dimension is shown through the following picture:

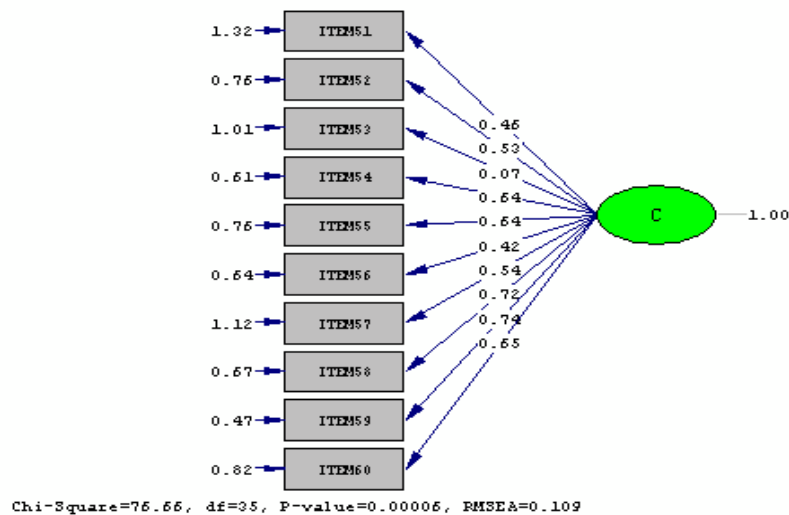


Figure 16. Estimation Result of Conventional Dimensional Measurement Model

The model on Figure 16 shows the size of *loading factor* in each indicator towards its latent variable. The estimated result of the parameter is presented in the following table 11:

Table 11. Evaluation of Factor Loading and Criteria Overall Measurement Model Fit Conventional (C) Dimension

Indicator	Factor Loading	t – value (Critical Ratio)
Item51	0.456	3.491
Item52	0.526	5.074
Item53	0.0673	0.609
Item54	0.644	6.527
Item55	0.637	5.936
Item56	0.416	4.476
Item57	0.540	4.401
Item58	0.722	6.860
Item59	0.742	7.832
Item60	0.649	5.849
Goodness of Fit Indices	Value	Decision
p value	0.00006	misfit
RMSEA	0.109	misfit

The result of statistic measurement on figure 16 and table 11 shows that all indicators of Conventional dimensional measurement have factor loading bigger than 0.5, except item51, item53, and item56 which have factor loading < 0.50 ; however, most of them are significant ($t > 1.96$) except for item53 with $t\text{-value} = 0.609$.

Observed from the result, the developed model is seen as misfit, which means zero hypothesis is rejected, which also means that the hypothesized model is dissimilar with the empirical data. Since the model is misfit ($p < 0.05$), it is important to do model modification. The

suggested modification is by looking at the modification indices, which provide information on the correlation between indicators of a latent construct. Based on the output, there is a correlation among the error variance on item52, item57, item58, and item59. The correlation between error variance informs that the indicators are strongly inter-correlated and explains one similar value related to Conventional dimension. If they are observed, item52, item57, item58, and item59 turn out having opposing sentence meaning, but the four of them are still the indicators of the similar dimension. If the four items are omitted from the model, the result could be seen in the following picture.

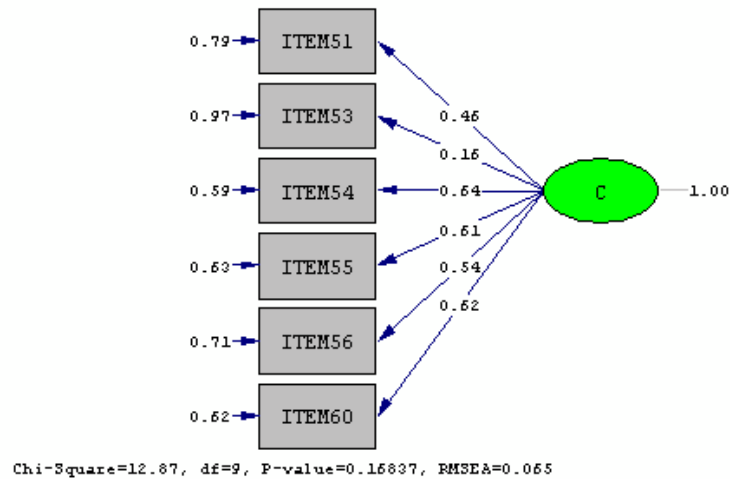


Figure 17. Result of *Standardized Solution* Conventional Dimensional Measurement Model (Revised Model)

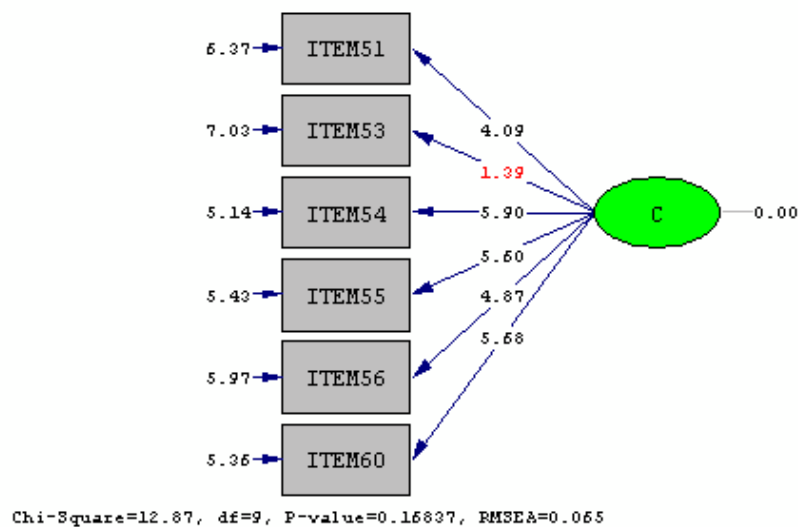


Figure 18. Result of *t-value* Conventional Dimensional Measurement Model (Revised Model)

The model (in the picture) shows the amount of *loading factor* of each indicator towards its respective latent variables. The amount of loading factor of each indicator is presented in table 12:

Table 12. Evaluation of Factor Loading and Criteria Overall Measurement Model Fit
Conventional (C) Dimension (Revised Model)

Indicator	Factor Loading	t – value (Critical Ratio)
Item51	0.46	4.09
Item53	0.16	1.09
Item54	0.64	5.90
Item55	0.61	5.60
Item56	0.54	4.87
Item60	0.62	5.68
Goodness of Fit Indices	Value	Decision
p value	0.16837	<i>Fit</i>
RMSEA	0.065	<i>Mediocre fit</i>

Figure 17, figure 18, and table 12 above present that by omitting six items, it could produce *good fits* since $p = 0.16837$ ($p > 0.05$) and mediocre fit on $RMSEA = 0.065$ ($RMSEA < 0.08$) (Darcey & Tracey in Louis, 2010). Generally, if the t value is observed carefully, all items will have t value bigger than 1.96 except for item53, which means that the indicator will conform to Conventional theory concept in RIASEC model. Therefore, if the model is to be used, it is better to omit item53 off the model.

V. CONCLUSION

This study examined the effectiveness of Holland Theory which was evidenced to be suitable as a basic theory for designing vocational interest instrument. This instrument was considered to valid and reliable in measuring the vocational interest of the students of junior high school. Subsequently, the data yielded by the instrument providing advice, suggestion, recommendation for the students particularly for choosing speciality of their next education level. It is expected that this instrument will be useful for the teacher and counselor in junior high school to give assistance to their students.

REFERENCES

- [1] D.G.J. Louis. The Development of an Interest Inventory Using Holland's RIASEC Typology. The International Journal of Educational and Psychological Assessment. April 2010, Vol. 4. 2010.
- [2] J. L. Holland. Making vocational choices: A theory of vocational personalities and work environments (3rd ed.). Odessa, FL: Psychological Assessment Resources. 1997.
- [3] M.U.A. Darcy & T.J.G. Tracey. Circumplex Structure of Holland's RIASEC Interests Across
- [4] Gender and Time. [Journal of Counseling Psychology, Volume 54, Issue 1](#), January 2007, Pages 17-31. 2007.
- [5] J.F. Jr. Hair, R.E. Anderson, R.L. Tatham & W.C. Black, Multivariate Data Analysis, (5th ed). Upper Saddle River, NJ: Prentice Hall. 1998.
- [6] K.S. Low & J. Rounds. Comprehensive Handbook of Personality and Psychopathology. New York: John Wiley & Sons, Inc. 2001.
- [7] W.S Winkel & H. Sri. Bimbingan dan Konseling di Institusi Pendidikan. Jakarta: PT. Grasindo. 2005.
- [8] J. Allik, K. Laidra, A. Realo & H. Pullmann. Personality Development from 12 to 18 Years of Age: Changes in Mean Levels and Structure of Traits. European Journal of Personality. Eur. J. Pers. 18: 445–462. 2004.
- [9] A.S. Al-Miskry, A.R. Bakar, & O. Mohamed. Gender Difference and Career Interest Among Undergraduates: Implication for Career Choices. European Journal of Scientific Research. Vol.26 No.3 (2009), pp.465-469. 2009.
- [10] P.I. Armstrong, W. Allison & J. Rounds. Development and Initial Validation of Brief Public Domain RIASEC marker Scales. Journal of Vocational Behavior. 73, 287-299. 2008.
- [11] K. Arrington. "Middle Grades Career Planning Programs". Journal of Career Development. Vol. 27.2. <http://jcd.sagepub.com>, diakses 15 Januari 2013. 2000.
- [12] S. Azwar. Penyusunan Skala Psikologi. Yogyakarta: Pustaka Pelajar. 2012.
- [13] Bullock, E.E., Andrews, L., Braud, J., & Reardon, R.C. Holland's Theory in an International Context: Applicability of RIASEC Structure and Assessments. Career Planning and Adult Development Journal. Winter 2009/2010; 25, 4; Proquest Education Journals, pg. 29. 2009.
- [14] W. R. Borg & M. D. Gall. Educational Research: An Introduction. New York & London: Longman. 1983.
- [15] K.Y. Chan. The Relation between Vocational Interests and the Motivation to Lead. Journal of Vocational Behavior. Vol 57, 226–245 (2000). 2000.
- [16] E. Cowger, I. Chauvin & M.J. Miller. An "Inverse" Validation of Holland's Theory. College Student Journal; September 2009; 43, 3; Proquest Education Journals, pg. 807. 2009.
- [17] R.V. Dawis. Handbook of Industrial and Organizational Psychology. Delhi: Jaico Publishing House. 1998.

-
- [18] G.P. De Bruin. The Relationship Between Personality Trait and Vocational Interest. *Journal of Industrial Psychology*, 28(1), 49-52. 2002.
 - [19] Depdiknas. Undang-undang No. 20 tahun 2003 tentang Sistem Pendidikan Nasional,
 - [20] Depdiknas. (2006) Panduan Penilaian Penjurusan Kenaikan Kelas dan Pindah Sekolah, Jakarta: Direktorat Pendidikan Menengah Umum. 2003.
 - [21] C.P. Deng, P.I. Armstrong, J. Rounds. The fit of Holland's RIASEC model to US Occupation. *Journal of Vocational Behavior*, 71 (2007) 1-22. 2009.
 - [22] D.S. Dimakakou, K. Mylonas, & K. Argyropoulou. Holland's Hexagonal Personality Model for Sample of Greek University Students. *International Journal of Educational and Vocational Guidance*. Vol: 8, 111-125. 2008.
 - [23] J. Farh, & F.T.L. Leong. Cross-Cultural Validity of Holland's Model in Hong Kong. *Journal of Vocational Behavior*. Volume 52, 425-440. 1998.
 - [24] K.A. Feldman, C.A. Ethingto & J.C. Smart. A Further Investigation of Major Field and Person- Environment Fit. *The Journal of Higher Education*. Nov/Dec 2001; 72, 6; Proquest Education Journals, pg. 670. 2001.
 - [25] J. Feist & G.J. Feist. *Theories of Personality*. New York: The McGram Hill Companies, Inc.
 - [26] Friedenberg, L. *Psychological Testing: Design, Analysis and Use*. New York: Schuster Publishing Company. 2006.
 - [27] A. Furnham, J. Zang & T. Chamorro-Premuzic. The Relationship Between Psychometric & Self Estimated, Intelligence, Creativity, Personality & Academic Achievement. *Imagination, Cognition, & Personality*. Vol. 25(2) 119-145, 2005-2006. 2005.
 - [28] R. K. Gable, *Instrument development in affective domain*. Boston: Kluwer-Nijhoff Publishing. 1986.
 - [29] L.R. Gay, *Educational Research: Competencies for analysis and application* (2nded.). Columbus: Charles E Merrill Publishing. 1981.
 - [30] G.D. Gottfredson & M.L. Johnstun. John Holland's Contributions: A Theory-Ridden Approach to Career Assistance. *The Career Development Quarterly*, December 2009; 58, 2; Proquest Education Journals, pg. 99. 2009.
 - [31] J.H. Greenhaus & G.A. Callanan, (ed). *Encyclopedia of Career Development*. Vol. 1. London: Sage Publication, Inc. 2006.
 - [32] D.T. Hall. *Career In and Out of Organizations*. Los Angeles: Sage Publications Inc. 2002.
 - [33] R. Hambleton & J. Rodgers. Item bias review. *PARE* Vol 4 Number 6. 1995.
 - [34] R. Hogan & R. Blake. John Holland's Vocational Typology and Personality Theory. *Journal of Vocational Behavior*, 55, 41-56. 1999.
 - [35] E.B. Hurlock, *Psikologi Perkembangan: Suatu Pendekatan. Sepanjang Rentang Kehidupan*. (Terjemahan Meitasari Tjandrasa & Muskichah Zarkasih). New York: McGraw-Hill. 2000.
 - [36] I. Irawati. Penjurusan Antara Minat Dan Obsesi Orang Tua. Artikel. <http://www.kabarindonesia.com/berita.php?pil=13&dn=20080704174933>. 2008.
 - [37] J.R.K. Kagaari. Evaluation of Effects of Vocational Choice and Practical Training on Students' Employability. *Journal of European Industrial Training*. Vol 31 No. 6, pp. 449-471. Emerald Group Publishing Limited. 2007.
 - [38] E.K. Kathryn & B. K. Lee You Do What Yaou Are: The Relationship Between the Scale of Creativity Attributes and Behavior and Vocational Interest. *Journal of Instructional Psychology*. Vol 36 .p. 79. 2009.
 - [39] I.S. Kirsch, & T. G. John. Construct Validity of Functional Reading Tests. *Journal of educational measurement*, Vol. 17, No. 2. 1980.
 - [40] E. Komandayhrini. Hubungan Self Efficacy dengan Kematangan dalam Memilih Karir Siswa Program Percepatan Belajar. *Jurnal Keberbakatan dan Kreativitas*, 2 (1), 1-12. 2008.
 - [41] L.M. Larson & F.H. Borge. Convergence of Vocational Interests and Personality: Examples in an Adolescent Gifted Sample. *Journal of Vocational Behavior*; 60, pg. 91-112. 2002.
 - [42] J. Leeson & R.C. Reardon. Using The RIASEC Schema in Expressed Vocational Assessment: An Australian Experience. *Career Planning and Adult Development Journal*; Winter 2009/2010; 25, 4; Proquest Education Journals, pg. 59. 2009.
 - [43] F.Y. Lubis. Pengembangan Alat Ukur Minat Untuk Pengembangan Karir Pada Lulusan Sekolah Menengah Atas. Laporan Penelitian. Tidak dipublikasikan. 2008.
 - [44] Djemari Mardapi. *Penilaian Pendidikan Karakter. Bahan Tulisan Penilaian Pendidikan Karakter UNY*. Tidak Diterbitkan. 2010.
 - [45] R.R. McCrae, & T. C. Paul Discriminant validity of NEO PIR facet scale. *Educational and psychology measurement*. 52,1, 229- 237. 1992.
 - [46] M.A. McDaniel, & A.F. Snell. Holland's Theory and Occupational Information. *Journal of Vocational Behavior*. 55, 74-85. 1999.
 - [47] M.J. Miller, T. P. Spinger, J.T. Tobacyk, & D. Wells. Congruency Between Occupational Daydreams and SDS Scores Among College Students. *College Student Journal*. 38, halaman. 57. 2004.
 - [48] M. J. Miller & T.A. Miller. Theoretical Application of Holland's Theory to Individual Decision-Making Styles: Implications for Career Counselors. *Journal of Employment Counseling*. Mar 2005; 42, 1; ProQuest Education Journals pg. 20. 2005.
 - [49] M.J. Miller, W.J. Scaggs & D. Wells. The Relevancy of Holland's Theory to a Nonprofessional Occupation. *Journal of Employment Counseling*. Jun 2006; 43, 2; ProQuest Education Journals pg. 62. 2006.
 - [50] R.O. Mueller. *Basic principles of structural equation modeling*. New York: Springer. 1996.

-
- [51] J.J. Muro & T. Kottman. *Guidance and Counseling in the Elementary and Middle Schoolas*. Madison: Wm C. Brown Com. Inc. 1995.
- [52] S. H. Osipow. *Theories of Career Development*. Englewood Cliffs: Prentice Hall, Inc. 1983.
- [53] W. Patton, M. McMahon. *Career Development and Systems Theory: A New Relationship*. Pacific Grove, CA: Brooks/ Cole Publishing. 1999.
- [54] R.C. Reardon, E.E. Bullock & K.E. Meyer. A Holland Perspective on the U.S. Workforce from 1960 to 2000. *The Career Development Quarterly*, 55(3), 262-275. 2007.
- [55] R.C. Reardon & J.G. Lenz. Holland's Theory and Career Assesment. *Journal of Vocational Behavior*, 55, 102-113. 1999.
- [56] P.J. Rottinghaus, A.R. Gaffey, F.H. Borgen & C.A. Ralston. Deverse Pathways of Psychological Majors: Vocational Interests, Self-Efficacy, and Intentions *The Career Development Quarterly*. 55, 1, pg. 85. 2006.
- [57] Santrock, *Life-Span Development: Perkembangan Masa Hidup*. Edisi Kelima. Jilid II. Jakarta: Erlangga. 2002.
- [58] Y. Satria. *Survey tentang Perencanaan Karir Siswa Kelas 3 SMP di Kota Solo*. Laporan Penelitian Reguler (tidak diterbitkan). Fakultas Psikologi UMS. 2013.
- [59] M.L. Savickas & A.R. Spokane. *Vocational Interest*. Palo Alto: Davies-Black Publishing. 1999.
- [60] M.L. Savickas & B.J. Taber. Individual Differences in RIASEC Profile Similarity Across Five Interest Inventories. *Journal of Measurement and Evaluation in Counseling and Development*, 38(4), 203-211. 2006.
- [61] L. Seligman. *Developmental career counseling and assessment* (2nded). Thousand Oaks: Sage Publishing. 1994.
- [62] G.J. Skulmoski, T. H. Francis & K. Jennifer. Delphi method for graduate research. *Journal of information technology education*. Volume 6. Diunduh pada tanggal 28 Desember 2011 dari <http://www.fepto.eu/storage/files/articole/Delphi%20method%20for%20Graduiate%20research.pdf>. 2007.
- [63] J.C. Smart & P.D. Umbach. Faculty and Academic Environments: Using Holland's Theory to Explore Differences in How Faculty Structure Undergraduate Courses. *Journal of College Student Development*; Mar/Apr; 48, 2; ProQuest Educatuonal Journals pg. 183. 2007.
- [64] J.C. Smart, C.A. Ethington, P.D. Umbach, and L.M. Rocconi. Faculty Emphases on Alternative Course-Specific Learning Outcomes in Holland's Model Environments: The Role of Environmentat Consistency. *Journal of Res High Education*. 50 : 483-501. 2009.
- [65] R.S. Sharf. *Applying Career Development Theory to Counseling* (4th Ed.). CA: Thomson- Brooks/Cole, Belmont. 2006.
- [66] R.E. Snow. Individual Differences and the Design of Educational Programs in *Journal of Psychology*, Vol 1, Dec. 1986, 22-35. 1986.
- [67] E. Susiati. *Hubungan antara Self Efficacy dengan Kematangan Karir Siswa*. Tesis magister, tidak diterbitkan, UPI Bandung. 2008.
- [68] J.R. Studer. *The Professional School Counselor: An Advocate for Student*. Belmont, CA: Thomson Brooks/Cole. 2005.
- [69] Taveira, Maria Do Ceu & M. Moreno, R. Luisa "Guidance Theory and Practice: The Status of Career Exploration". *British Journal of Guidance and Counseling*. Vol. 31.2. <http://proquest.-umi.com/pqdweb>, diakses 15 Januari 2013. 2003.
- [70] W.U. Tien. Vocational interest and career maturity of Male High School Students talented in Mathematics and Science. *Journal of Vocational Behavior*. National Science Council Production. 10(3), 137-143. 2009.
- [71] D.M. Tokar, A.R. Fisher & L.M. Subich. "Personality and Vocational Behavior: A Selected Review of the Literature", 1993 – 1997. *Journal of Vocational Behavior*. 53. 115 – 153. <http://jcd.sagepub.com>, diakses 25 Januari 2013. 1998.
- [72] K.D. Toomey, E.M. Levinson & E.J. Palmer. A Test of Holland's Theory of Vocational Personalities and Work Environments. *Journal of Employment Counseling*. Vol. 46, p. 82. 2009. [71] Unesco. *Handbook on Career Counselling*. Published in 2002 by the United Nations Educational, Scientific and Cultural Organization 7, place de Fontenoy, 75352 Paris 07 SP. 2002.
- [73] M. T. Wicaksono. *Pengantar Bimbingan dan Konseling Karier*. Jakarta: Bumi Aksara. 2009. [73] Witko, Kim, Bernes, Kerry B., Magnusson, Kris, & Bardick, Angela D. 2005. "Senior High School Career Planning: What Students Want". *Journal of Educational Enquiry*. Vol. 6.1. <http://proquest.umi.com/pqdweb>, diakses 21 Januari 2013. 2005.
- [74] I. Wijayanti. *Penjurusan, Antara Minat dan Obsesi Orang Tua*, Artikel. www.smam6paciran.com. 2010.
- [75] C.S. Wong & P.M. Wong. Validation of The Measurement Scale and the Vocational Orientation Model in Four China Societies. *Journal of College Student Development*. Vol. 16, p. 165. 2009.

Self-Assessment of Teachers of Mathematics Vocational High School in Yogyakarta city on the Performance Post-Certification

Zuli Nuraeni

STKIP of MuhammadiyahKuningan
zulinuraeni.wsb@gmail.com

Abstract - This study aims to describe the post-certification performance of mathematics teachers of vocational high schools in: (1) planning the lesson, (2) the learning implementation, (3) the learning assessment, and (4) the professional development. This was a descriptive study with quantitative approach. The research subjects were all mathematics teachers of vocational high schools in a regency who have passed the teachers certification. The instrument to collect the data consisted of teachers' self-assessment questionnaires. Teacher self-assessment questionnaire was used to uncover performance of mathematics teachers of vocational high schools in planning the lesson, the learning implementation, the assessment learning and the professional development after certification. The data analysis used in this study is an interactive analysis which consists of three components: data reduction, data presentation and conclusion. To get conclusion the data categorized performance trends into 5 groups: Very Good, Good, Fair, Poor, and Very Poor. The results showed that: (1) Performance math teacher Vocational High School in Yogyakarta City post-certification in lesson planning, including the excellent category; (2) Performance math teacher Vocational High School in Yogyakarta City post-certification in the implementation of learning including the very good categories; (3) Performance math teacher Vocational High School in Yogyakarta City post-certification in the assessment of learning, including the very good category; and (4) Performance math teacher Vocational High School in Yogyakarta City post-certification in professional development category is very less and needs to be improved.

Keywords: self-assessment, performance of teacher, certification

I. INTRODUCTION

Education have an important role in the formation and development of human resources. Sihombing&Sihombing (2011, p.101) say that the quality of education is largely determined by the quality of the teachers as actor. So, the teachers are required to be a professional and have high performance in order to improve the quality of the young generation so that Indonesia could become a intelligent nation and able to face the future challenges.

The research results from Pearson Learning Curve in 2013 against quality of Indonesia's education mapping shows that Indonesia is in a position 40 of the 40 countries studied (Baswedan, 2014). In addition, the results of teachers competency test in 2012 to 460,000 teachers showed the average value of the competency test for teachers only 44.5 of the standard value of the expected average of 70 (Baswedan, 2014). The low quality of education in Indonesia causes vary, such as: students' achievement of learners were low, quality of infrastructure were poor, quality of teachers were poor, teachers' welfare were low, educational equity opportunities were low, relevance of education to the needs were poor, and the cost of education were high (Setyaji, 2011). Further, Utami (2015, p.473) mentions the other problems of education in Indonesia are as follows: high number of unqualified teachers, the uneven teacher distribution and over supply contract teachers.

To follow the teacher certification, the teachers must have competencies. Muzenda (2013, p.7) says that competence is regarded as a multidimensional construct teaching which encompasses numerous interconnected elements towards transformation of knowledge to learners. Meanwhile, Bhargava & Pathy (2011, p.77) say that competencies are specific and demonstrable characteristics or attributes inevitable for teaching professionals to create a convincing and learner friendly environment. Teachers competencies as knowledge and skills of teachers required for effective and quality education at higher education level. These include a set of teaching skills that a teacher at the tertiary level needs to possess, in order to become effective teacher and these are pedagogical skills, management and assessment skills, and research skills (Aziz & Akhtar, 2014, p.122).

Khaeruniah gives the statement (2013, p.108) that a teacher's competence is the ability of a teacher to realize the planned educational aim. Performance indicators referring to the performance appraisal indirectly that the things that seem only an indication of performance, so that the shape tends to qualitative or can't be calculated such as increased, accuracy, velocity, level, effectiveness, and others (Moeheriono, 2012, p.32-33). To determine whether a teacher's performance is quite optimal or not can be seen from various indicators. With a performance assessment will help teachers in identifying a better job so that the teachers will run the learning process as effective as possible for students' progress and education (Barnawi & Arifin, 2012, p.25).

Schacter (2000, p.14) splits indicator of teachers' performance in three parts, are: (1) the skills, knowledge, and responsibility of the teacher; (2) students' achievement at grade level; and (3) achievement in school. Meanwhile, Usman (2006, p.10-19) suggests there are three indicators of the teachers' performance. First, the ability to planning the lesson, that: (1) control of the outlines of the organization of education; (2) adjusting the analysis of the subject matter; (3) collate semester program; and (4) develop a program or learning. Second, the ability to carry out the teaching and learning activities, are: (1) pre instructional; (2) the instructional stage; and (3) evaluation stage and no further. Third, the ability to evaluate, are: (1) the normative evaluation; (2) formative evaluation; (3) report the results of the evaluation, (4) the implementation of the improvement and enrichment programs. Ditjen PMPTK, Ministry of National Education (2008, p.22) suggests indicators on teachers' performance assessment conducted on three classroom activities, that are planning, execution and the learning assessment.

It can be concluded that teachers' performance is the result that can be achieved by teachers in carrying out the duties which they are responsible is based on skills, experience, and sincerity to work within a certain time frame. Teachers' performance did not materialize for granted, but it is influenced by certain factors, both internal factors and external factors. Factors that affect performance include mental attitude (work motivation, work ethic), educations, skills, leadership management, income levels, salaries and health, social security, labor climate, infrastructure, technology and outstanding opportunities (Asf& Mustafa, 2013, p.160).

In this study, the teachers' performance in general are intended as a performance in the implementation of a teaching assignment with the four indicators, are as follows: (1) performance in planning the lesson; (2) performance in the implementation of learning; (3) performance in the learning assessment, and (4) the performance in professional development.

Government efforts to improve the quality of human resources have been carried out. Not only by improving the quality of teachers, but also improve the quality of students' graduates. Learners are equipped with the competence and expertise to be ready to work so that she could survive for life. Currently vocational education become a priority of government, created vocational education graduates who are ready to print the job and have the skills to work. Vocational education is built and developed with the needs and situation of the workforce to meet the growing market demands (Sriadi, 2011, p.32).

Most of mathematics teachers at vocational high schools have not been able to package the matter of mathematics well, the contextual, flexible and fun, so that vocational students' achievement can't be maximized. Research results from Shadiq (2006, p.10-11) identifies the difficulties encountered some mathematics teachers of vocational high schools at mathematics learning such as teachers tough for: (1) understand and apply the philosophy of constructivism in contextual learning; (2) encourage students to formulate their own problems in conducting discovery (inquiry); (3) encourage and assist students to make connections between the

knowledge that have been had by the newly learned knowledge; (4) design issues categorized as a problem; (5) get a reference books that are relevant; (6) familiarize students to express ideas; and (7) the teachers were difficult in guiding students to formulate a conjecture of existing data.

II. RESEARCH METHODS

This study was included in the descriptive research with quantitative approach. Research conducted at several vocational high schools in Yogyakarta city which have a certified mathematics teachers educators. The subjects in this study are all of mathematics teachers of vocational high schools in Yogyakarta city which have passed the certification.

The instrument to collect the data consisted of teachers' self-assessment questionnaires, Teacher self-assessment questionnaire was used to uncover performance of mathematics teachers of vocational high schools in planning the lesson, the learning implementation, the assessment learning and the professional development after certification. It's shaped scale inventory questionnaire using Likert scale model of a modified form of multiple choice. Multiple choices consisting of 4 answer choices, where each answer have been determined score.

A. Validity and Reliability

The instruments validity in this study was proved by content validity and construct validity. Proving the validity of the content by using expert judgment then proved by Aiken's validity (Retnawati, 2014, p.3). Get the proof validity of the construct using analysis factor. The results of the both proving validity that prove that the instruments are valid because it has V value of more than 0.5. The analysis factor to 8 items of questionnaire of performance in planning the lesson got the KMO value is $0.784 > 0.5$ with significance 0.00 and there is a component that have Eigen's value > 1 and variance total is 60.277%. The analysis factor to 9 items of questionnaire of performance in implementation of learning got the KMO value is $0.839 > 0.5$ with significance 0.00 and there are 2 components that have Eigen's value > 1 and variance total is 66.751%. The analysis factor to 9 items of questionnaire of performance in assessment of learning got the KMO value is $0.821 > 0.5$ with significance 0.00 and there are 2 components that have Eigen's value > 1 and variance total is 65.972%. The analysis factor to 13 items of questionnaire of performance in development professional didn't get the KMO value cause the data have negative definite. This was because many items that is unfilled, but there are 4 components that have Eigen's value > 1 and variance total is 76.448%.

Estimate of reliability instrument of the mathematics teachers of vocational high schools using by Alpha Cronbach. Having obtained the coefficient of reliability of the instrument, then calculate the Standard Error Measurement (SEM), to understand the size of the measurement error for assessment procedures each instrument. Estimate of reliability and SEM as shown in Table 1.

Table 1. Estimate of Reliability Instrument

Instrument		Reliability	SEM
Teacher's self-assessment questionnaire	Planning the lesson	0.904	2.087
	Implementation of learning	0.891	2.062
	Assessment of learning	0.868	2.025
	Professional Development	0.809	4.621

Data analysis techniques used in this research is descriptive quantitative techniques of data analysis. The effectiveness of performance of mathematics teachers of vocational high schools is determined based on the level of propensity to perform categorization of variables. Trend level of the post-certification performance of mathematics teachers of vocational high schools is divided into 5 categories as shown in Table 2.

Table 2. Categories of The Post-Certification Performance of Mathematics teachers of Vocational High Schools

Formulas	Category
$X > \bar{X}_i + 1,8 \times sb_i$	Very Good
$\bar{X}_i + 0,6 \times sb_i < X \leq \bar{X}_i + 1,8 \times sb_i$	Good
$\bar{X}_i - 0,6 \times sb_i < X \leq \bar{X}_i + 0,6 \times sb_i$	Fair
$\bar{X}_i - 1,8 \times sb_i < X \leq \bar{X}_i - 0,6 \times sb_i$	Poor
$X \leq \bar{X}_i - 1,8 \times sb_i$	Very Poor

Specification:

X : score of respondents or actual score

\bar{X}_i : a mean score of ideal

SB_i : ideal standard deviation

III. RESULTS AND DISCUSSION

A. *Performance of Mathematics Teachers of Vocational High Schools in Planning the Lesson*

Based on the results from self-assessment questionnaire of mathematics teachers of vocational high schools in Yogyakarta city after certification in planning the lesson, the actual mean score 26.11 (good category), the ideal mean score 20, the ideal standard deviation 4, the ideal maximum score 32, and the ideal minimum score 8. Percentage of the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in planning the lesson according to the teachers' self-assessment are 53% were in the very good category, 30% were in the good category, 14% were in the fair category, and 3% were in the poor category. Thus it can be said that the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in planning the lesson, including the good category.

B. *Performance of Mathematics Teachers of Vocational High Schools in The Learning Implementation*

Based on the results from self-assessment questionnaire of mathematics teachers of vocational high schools in Yogyakarta city after certification in the learning implementation, the actual mean score 31.32 (very good category), the ideal mean score 22.5, the ideal standard deviation 4.5, the ideal maximum score 36, and the ideal minimum score 9. Its percentage are 62% were in the very good category, 33% were in the good category, and 5% were in the fair category. It can be concluded that the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in the learning implementation is in the very good category.

C. *Performance of Mathematics Teachers of Vocational High Schools in The Learning Assessment*

Based on the results from self-assessment questionnaire of mathematics teachers of vocational high schools in Yogyakarta city after certification in the learning assessment, the actual mean score 30.84 (good category), the ideal mean score 22.5, the ideal standard deviation 4.5, the ideal maximum score 36, and the ideal minimum score 9. The percentage is as much as 54% were in the very good category, 43% were in the good category, and 3% were in the fair category. It can be concluded that the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in the learning assessment including the good category.

D. *Performance of Mathematics Teachers of Vocational High Schools in the Professional Development*

Based on the results from self-assessment questionnaire of mathematics teachers of vocational high schools in Yogyakarta city after certification in the professional development, the actual mean score 11.46 (very poor category), the mean ideal score 32.5, the ideal standard deviation 10.83, the ideal maximum score 65, and the ideal minimum score 0. The percentage are 57% were in the poor category, 38% were in the very poor category, and 5% were in the good category. Thus it can be said that the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in professional development including the category of very poor and needs to be improved.

Here the overall profile of the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in the planning, implementing, and the learning assessment and in the professional development.

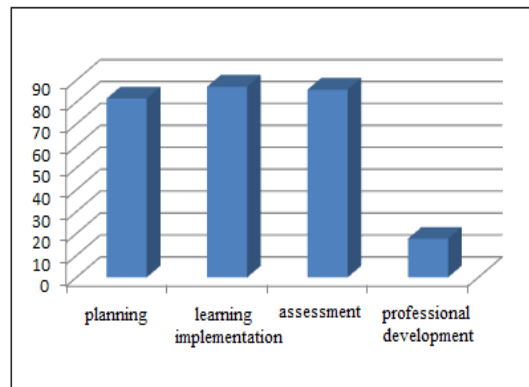


Figure 1. Appraisal of the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city

IV. DISCUSSION

Based on the description of the results of research obtained information that the post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in planning the lesson based on self-assessment is in the good category, where 53% were in the very good category, 30% were in the good category, 14% were in the fair category, and 3% were in the poor category. Over all, mathematics teachers of vocational high schools in Yogyakarta city after certification have good performance in planning the lesson aspects. This was confirmed by the results of Khodijah's research (2013, p.95) regarding the post-certification performance of Religion teachers in the province of South Sumatra which revealed that the performance of teachers in planning the lesson in the good category. The results of Palupi's research (2011, p.187) of regarding the post-certification performance of Science teachers in the province of Yogyakarta also revealed that the performance of teachers in planning the lesson in the good category.

Performance mathematics teachers of vocational high schools in Yogyakarta city after certification in the learning implementation based on self-assessment showed the very good category, in which 62% were in the very good category, 33% were in the good category, and 5% were in the fair category. Over all, mathematics teachers of vocational high schools in Yogyakarta city after certification have very good performance in the learning implementation aspect. This was confirmed by the results of Kartowagiran's research (2011, p.467) which states that the post-certification performance of teachers in Sleman regency in the implementation aspect of learning undertaken by teachers including excellent category.

Performance of mathematics teachers of vocational high schools in Yogyakarta city after certification in the learning assessment based on self-assessment showed is in the good category, where 54% were in the very good category, 43% were in the good category, and 3% were in the fair category. Over all, mathematics teachers of vocational high schools in Yogyakarta city after certification have good performance in the learning assessment aspect. This was confirmed by the results of Khodijah's research (2011, p.467) which states that the post-certification performance of religion teachers in the province of South Sumatra which revealed that the performance of teachers in the learning assessment in the good category

Performance of mathematics teachers of vocational high schools in Yogyakarta city after certification in professional development based on self-assessment showed is still not good, in which more than half (57%) were in the poor category, 38% were in the very poor category, and 5% were in the good category of the number of teachers who are already certified, the performance mathematics teachers in professional development is in the poor category and half more (50%) is in the very poor category. Over all, mathematics teachers of vocational high schools in Yogyakarta city after certification have very poor performance in the professional development aspect.

V. CONCLUSIONS

Based on the analysis of data and discussion, this study can be concluded that: (1) The post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in planning the lesson is in the good category; (2) The post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in the learning implementation is in the very good category; (3) The post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in the learning assessment is in the good category; (4) The post-certification performance of mathematics teachers of vocational high schools in Yogyakarta city in the professional development is in very poor category and needs to be improved.

REFERENCES

- [1] [1] Asf, J., & Mustofa, S. (2013). *Supervisi pendidikan*. Yogyakarta: Ar-Ruzz Media.
- [2] [2] Aziz, F., & Akhtar, M.M.S. (2014). Impact of training on teachers competencies at higher education level in Pakistan. *International Refereed Research Journal*, 1, 121-128.
- [3] [3] Bhargava, A., & Pathy, M. (2011). Perception of student teachers about teaching competencies. *American International Journal of Contemporary Research*, 1(1), 77-81.
- [4] [4] Baswedan, A.R. (2014). Gawat darurat pendidikan di Indonesia. Materi presentasi dalam Silaturahmi Kementerian dengan Kepala Dinas di Jakarta 1 Desember 2014. Diambil pada tanggal 15 April 2015, dari <http://www.kemdiknas.go.id/kemdikbud/files/Paparan%20Menteri%20-%20Kadisdik%2014120%20Low20.v.0.pdf>.
- [5] [5] Barnawi, & Arifin, M. (2012). *Kinerja guru profesional*. Yogyakarta: Ar-Ruzz Media.
- [6] [6] Kartowagiran, B. (2011). Kinerja guru profesional (Guru pasca sertifikasi). *Cakrawala Pendidikan*. Tahun XXX, No. 3, 463-473.
- [7] [7] Khodijah, N. (2013). Kinerja guru madrasah dan guru pendidikan agama Islam pasca sertifikasi di Sumatera Selatan. *Cakrawala Pendidikan*. Tahun XXXII, No. 1, 91-102.
- [8] [8] Moeheriono. (2012). *Indikator Kinerja Utama (IKU)*, Jakarta: Raja Grafindo Persada.
- [9] [9] Muzenda, A. (2013). Lecturers' competences and students' academic performance. *International Journal of Humanities and Social Science Invention*, 3(1), 6-13.
- [10] [10] Palupi, R. (2011). Kinerja guru rumpun IPA SMA (biologi, kimia, dan fisika) se-propinsi Daerah Istimewa Yogyakarta pascasertifikasi. *Jurnal Inovasi Pendidikan IPA*, 1(1), 187-201.
- [11] [11] Retnawati, H. (2014). Membuktikan validitas instrumen dalam pengukuran. Downloaded on Juli 1st 2015 from <http://evaluation-edu.com/wp-content/uploads/2014/10/2-Validitas-heri-Retnawati-uny.pdf>
- [12] [12] Schacter, J. (2010). Teacher performance based accountability: Why what and how. Downloaded on September 10th 2014 from http://www.mff.org/performance_assessment.pdf
- [13] [13] Setyaji, A.M. (2013). Pelajari dulu wajah pendidikan. Downloaded on September 10th 2014 from <http://www.psmkgi.org/read/875658/pelajari-dulu-wajah-pendidikan.html>
- [14] [14] Shadiq, F. (2008). Identifikasi kesulitan guru matematika SMK pada pembelajaran matematika yang mengacu pada permendiknas no. 22 tahun 2006.
- [15] [15] Sihombing, D., & Sihombing, M. (2011). Analisis korelasi motivasi kerja dengan kinerja guru. *Jurnal Darma Agung*, J-DA, 101-104.
- [16] [16] Usman, M.U. (2006). *Menjadi guru profesional*. Bandung: Remaja Rosdakarya.
- [17] [17] Utami, IG.A.L.P. (2015). Teacher certification program in Indonesia: Problems and recommendation for the betterment of the program. *International Journal of English and Education*. 4(2), 47-489.

The measurement Model of Historical Consciousness

Aisiah

Universitas Negeri Padang

Aisiah.unp@gmail.com

Abstract- The study was to develop a measurement model of historical consciousness through a research and development model that had been adopted from the Plomp model. The historical consciousness was measured from four components i.e. knowledge of historical events, understanding of historical research method, meaning of historical events and usefulness of history. The development procedures in the model included a preliminary investigation in the form of literary study about the constructs of historical consciousness. In the design stage, the researcher designed a conceptual model and a hypothetical model about the historical consciousness. Then, the researcher performed a test construction namely assembling the test instrument for measuring the historical consciousness. Eventually, the researcher performed the test, the evaluation and the revision. The test in the study referred to the experiment toward the instrument, while the evaluation in the study referred to the efforts of identifying the obstacles that the participants encountered within the experiment toward the test in order to revise the test instrument. The experiment of the test involved history teacher-candidate university students in the Yogyakarta State University and the Padang State University. For the data gathering technique and instrument, the researcher implemented the associative multiple choice design. For the construct analysis, the researcher implemented the confirmatory factor analysis by means of Lisrel 8.80 program. The results of the analysis showed that the χ^2 value had been equal to 121.98, the p-value had been equal to 0.11, the RMSEA value had been equal to 0.043. In other words, the measurement model of historical consciousness that had been developed was supported by the empirical data.

Keywords: *historical consciousness, measurement model*

I. INTRODUCTION

This article reviews the measurement model of historical consciousness. The intended measurement model is an effort to make a measurement model of a latent variable (historical consciousness) through components or indicators that will be presented in the form of path diagram. The concept of historical consciousness is defined as a condition or a reasoning process in which people recall the meaning or the usefulness of history. The constructs or the components of historical consciousness are derived from the ideas of the Indonesian historians like Soedjatmoko, Ruslan Abdulgani, Sartono Kartodirdjo and more.

Every citizen should develop the historical consciousness in their nation and state life [1]. The effort of developing the historical consciousness among the generation of a nation might be pursued by means of historical education (learning). Kartodirdjo [2] asserts that historical subject has a socio-cultural function that is to encourage the historical consciousness. The historical consciousness is a key concept that has been very important [3] and significant [4] in historical didactic. People who study history has the ability to compare the difference of periods, cultures and social systems [5]. This ability is the manifestation of an individual's historical consciousness.

The historical learning recently has not been successful in developing the historical consciousness among the young generation. The condition is definitely concerning. In the university, for example, there are some students who have not understood and even comprehended the important meanings of their nation's history [6]. The emphasis on the factual knowledge is certainly "dry" and does not uncover much of students' understanding toward the exemplary values that will be studied or be researched in the final assignment. According to Mardapi [7], the learning quality might be viewed from the assessment results. Both aspects are related from one to another and there should be continuous improvement efforts.

The measurement of historical consciousness level for the history teacher-candidates become an urgent demand regarding the fact that these history teacher candidate will be the history teachers who should inculcate the historical consciousness among the learning participants. Therefore, this study is intended to develop a measurement model of historical consciousness.

A. The Constructs of Historical Consciousness

The researcher formulates the constructs of historical consciousness from the ideas of several Indonesian historians regarding the concept of historical consciousness and their ideas are presented in Table 1.

Table 1. The Historians' Views Regarding the Concept of Historical Consciousness

Historians	Definition of Historical Consciousness	Conclusion Regarding Construct
Ruslan Abdulgani	Historical consciousness has been a mental attitude (strength) that covers the knowledge of historical facts and their causality, the logic of history and the improvement of conscience by wisdom and intelligence for reflecting from the past.	1. Knowledge of Historical Events (<i>Pengetahuan Peristiwa Sejarah</i> , PPS) 2. Understanding of Historical Research Method (<i>Pemahaman Metode Penelitian Sejarah</i> , PMPS) 3. Meaning of Historical Events (<i>Pemaknaan Peristiwa Sejarah</i> , MPS) 4. Usefulness of History (<i>Kegunaan Sejarah</i> , GS)
Sartono Kartodirdjo	Historical consciousness will be improved by possessing the historical knowledge, the historical mindedness and by being able to imagine the situation of past history, cultural atmosphere, sentiment, idea, mentality, life style etc.	
Soedjatmoko	Historical consciousness has been a mental attitude and a manner to put one self in front of the truth and the social reality in the perspective of present, past and future.	
Adrian Bernard Lopian	Historical consciousness can't be separated from clarification, that is a historical investigation that entails the elementary aspects such as who, what, when, where and why, the historical impression and function in the education and the controversial aspects.	
R. Z. Leirissa	Historical consciousness is an understanding toward the essence of historical study.	
Ayatroehadi	Historical consciousness includes insight regarding history, the ideas within the historical insight, the theoretical and methodological foundation of historical study and the oral/written review regarding history.	

The constructs of historical consciousness are derived from the definition of historical consciousness concept according the ideas of the historians that have been presented in the table 1 above. The operationalization of the historical consciousness concept is adopted from Greenberg's idea [8] who states, "historical consciousness as a conceptual system comprising interactive elements which allows comperehension of temporal or historical experience and individual placement in time/history." In other words, historical consciousness as a conceptual system consists of some aspects that form the historical consciousness. Therefore, the researcher designs a conceptual system of historial consciousness that consists of four components and will be the hypothetical constructs in developing the measurement model of historical consciousness (Figure 1).

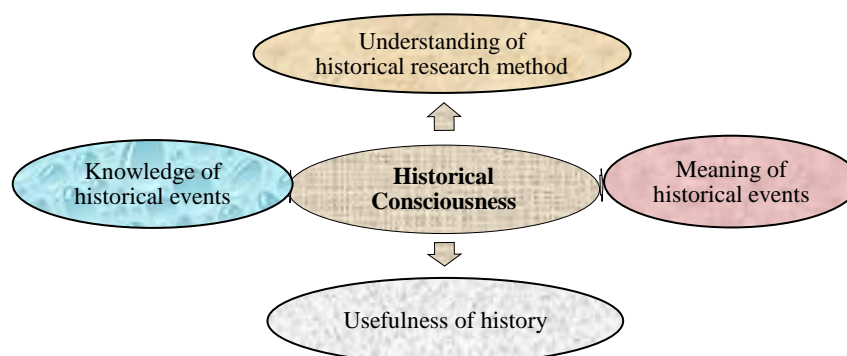


Figure 1.

Figure 1 is a conceptual system in the form of the four constructs that will establish the historical consciousness. The appearance of historical consciousness starts from the knowledge upon the historical facts [9] and the interrelatedness among the historical facts [10]. The historical knowledge can't be separated from the process of historical study or historical research method implementation. The historical knowledge is proved by the robustness of research findings [11]. The results of historical studies might strengthen the historical consciousness and even the historical consciousness might be the foundation in differentiating the fact and the fantasy [12] in the process of historical reconstruction.

The meaning of historical events lies in the significance that people give to the historical events [13]. The significance of historical events is shown by the values that has been contained in the values that individuals held in the past [14] and in the impacts of the historical events. By recalling the past an individual might act better in the upcoming future [15]. The expectation is that an individual will not commit the same mistake that took place or that had been experienced in the past for the sake of the upcoming future.

B. Knowledge of Historical Events

Knowledge of historical events is the knowledge about what (event) has occurred in the past of human history or the knowledge about the historical facts and processes [16]. The knowledge of historical events is not lied in what aspects that might inform the future; instead, the knowledge of historical events is lied in what aspects that might inform the past [17]. The knowledge of historical events might be measured through what has been recalled regarding the facts that have been learned [18].

Intellectual curiosity regarding the matters of the past is one of the reasons why people learn and study history [19]. Historical knowledge is one of the historical understanding elements [20]. The historical understanding includes an understanding toward causality [21]. The students consume and produce information through texts and develop their skills in the interpretation and the construction of historical knowledge and ideas [22]. The students might put themselves into the consumers and the producers of historical knowledge when they are studying history.

C. Understanding of Historical Research Method

Historical research method refers to the use of a sequence of scientific procedures to verify the historical evidence or sources [23]. These procedures include the topic selection, the heuristic, the internal and external criticism, the analysis and interpretation and the presentation in the form of historical written [24]. The topic selection should be in accordance with the researcher's interest.

The historical sources consists of primary sources and secondary sources. After the historical sources (documents) have been found, there are two aspects that should be investigates, that is the authenticity and the credibility of the sources [25]. This process is called as source criticism. The source criticism consists of internal and external criticism. The external criticism refers to the efforts of proving the source authenticity by investigating the physical sources or testing the external aspects of the historical sources [26]. Interpretation covers analysis and synthesis. Analysis refers to elaboration, while synthesis refers to unification [24]. After the researcher has found the data, the data then will be analyzed and, therefore, the researcher will find the historical facts. People might have different opinions in the analysis and the synthesis.

D. Meaning of Historical Events

The meaning of historical events for people, objects and events depend on the value implementation of certain perspectives [27]. The historical meaning is shown by the historical significance. The effort of training the capacity in building meaning is a matter that exceeds the simple knowledge-based content [28]. People might define or find the important meaning and significance of historical events by understanding the complexity of the events.

History does not only share past memories but also understand the meaning of these past memories. Questions regarding the historical meaning are the ones that live forever and that will always be raised by human beings [9]. History might be said as having historical meaning if the history might deliver human to the discovery of future points. The human conscience becomes the basis of self-awareness from the life experiences in which the historical meaning is uncovered [27]. Past experiences become a useful guidance for encountering the future.

E. Usefulness of History

History has multiple usefulness. The usefulness values of history might be viewed from the theoretical and the practical aspects. The theoretical usefulness are related to the tendency of learning past events for the sake of intellectual-academic needs (scientific interest) of the history [10]. All historical knowledge are based on the practical needs of human beings [29]. The practical usefulness of historical learning might be viewed from the educational, instructional, inspirational and recreational aspects.

History is useful for the educational purpose and giving good lesson. By learning history, people might find many educational examples in the form of moral actions and attitudes that should be attended and that should be avoided. History is also useful for serving as learning materials [30]. By reading multiple historical studies (autobiography and biography) individual might attain inspirations where they want to go [24]. People might see the past to find the solution for the recent problems [25]. By learning history people will be creative in encountering the challenges of the century. History is a set of experiences that become the basis for projecting the future and for predicting the upcoming events [19]. Past experiences also be the basis for anticipating every single possibility that might occur in the future.

F. Measurement Model of Historical Consciousness

Measurement is the process of defining numbers in a systematic way for stating the condition of an individual or an object [31]. Individual condition which is measured in the education domain is usually related to the learning results. In historical learning, the measurement of learning results might be directed to the historical consciousness. A measurement model shows the relationship between one observed variable (indicator or observed response) to another that becomes the basis of latent variables [32] state that measurement model describes how good the indicators can be implemented as measuring instrument of latent variable (an abstract object) such as knowledge, behaviors and attitudes of human beings.

Hendryadi [33] also asserts that measurement model has been an effort of creating measurement modelling from the latent variables through dimensions or indicators. Kusnendi [34] concretely states that measurement model as the operationalization of latent variable or research constructs becomes the measurable indicators that will be formulated into certain path diagram. Khine [32] asserts as well that in a wider sense measurement model determines how a theory will be operationalized as latent variables and observed variables.

Researchers design measurement models by implementing the theoretical knowledge or the empirical study and then hypothesize the relationship pattern between the observed variables and the latent variables. The theory plays an important role in the construction of measurement model [32]. The results of theoretical review regarding the latent variable (historical consciousness) show that the historical consciousness has been established by four aspects (figure 1). The indicators of each dimensions will be presented in Table 2.

Table 2. Components and Indicators of Historical Consciousness

Aspect	Indicator
Knowledge of Historical Events (<i>Pengetahuan Peristiwa Sejarah</i> , PPS)	Understanding the facts of historical events, including: what (event), who (figure), when (time), where (place) and why (cause)
Understanding of Historical Research Method (<i>Pemahaman Metode Penelitian Sejarah</i> , PMPS)	Identifying the procedures of conducting historical study, including: heuristics, criticism, verification, interpretation and historiography
Meaning of Historical Events (<i>Pemaknaan Peristiwa Sejarah</i> , MPS)	Finding the positive impact, the negative impact and the positive values of historical events
Usefulness of History (<i>Kegunaan Sejarah</i> , GS)	Identifying the usefulness of history theoreticall and practically (instructional, educational, inspirational, recreational and predictional usefulness)

The theoretical constructs of historical consciousness that contains the components and the indicators of historical consciousness become the basis of historical consciousness hypothetical masurement model development and the researcher formulates the hypothetical model development of historical consciousness in Figure2 as follows

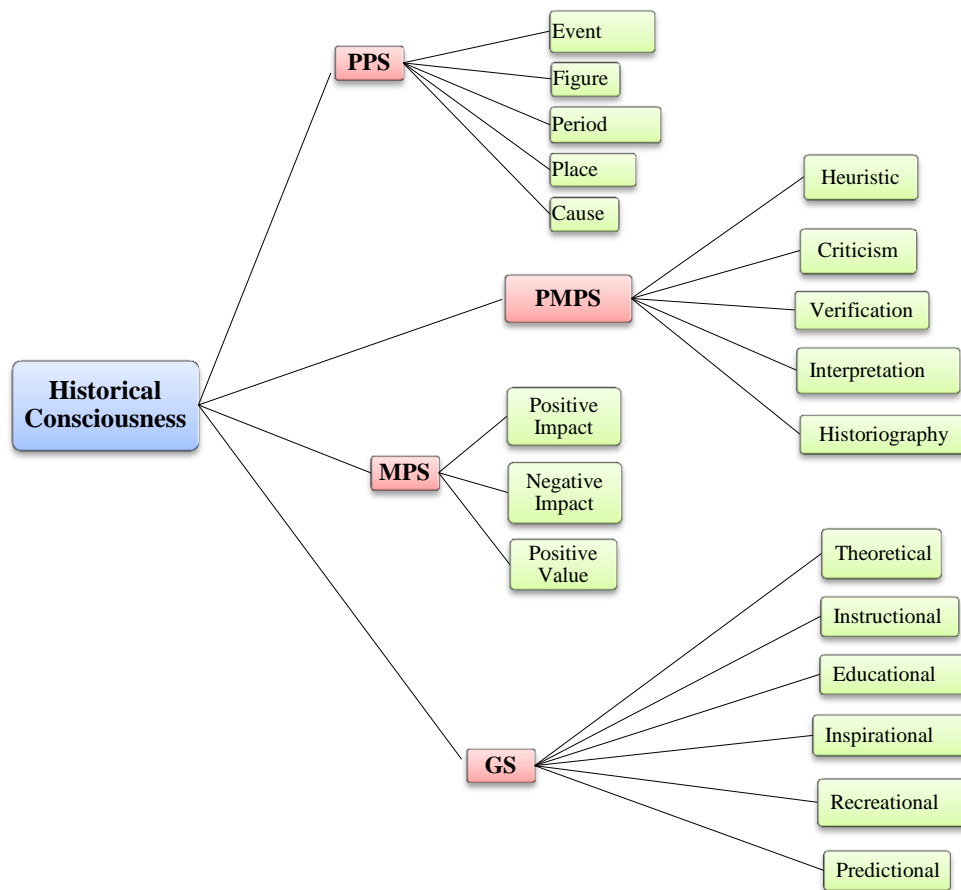


Figure 2. The Hypothetical Measurement Model of Historical Consciousness from the Results of Theoretical Review

II. METHOD

The study implemented research and development method that had been adapted from the Plomp model [35]. The procedures of development included preliminary investigation (literary study), the design (formulate constructs and hypothetical diagram of historical consciousness measurement model), construction (assembling instrument) and test (empirical testing) for estimating validity and reliability of the constructs instrument and measurement model of historical consciousness.

The data gathering was conducted by empirical testing the instrument in the form of associative multiple choices for measuring the historical consciousness. The empirical testing was conducted by testing the instrument toward the history teacher candidates as the research subjects at Yogyakarta State University and Padang State University. The selection of the subjects was conducted in a stratified sampling by involving the first, the third and the fifth history teacher candidates 2014/2015 Academic Year. The total number of the subjects was 190 people.

Table 3. Details of Research Subjects

University	Yogyakarta State University		Padang State University			Total
Semester	I	V	I	III	V	
Total	45	27	28	48	42	190
	72		118			

The data analyzed with confirmatory factor analysis by means of Lisrel 8.80 software. The goal was to test the validity and reliability as well as the fitness of the measurement model that had been formulated. Khine [32] asserted that confirmatory factor analysis had been frequently

implemented in the measurement model testing. The validity and reliability of measurement model would be shown by the construct validity and reliability of the measuring tool that would be used. The construct validity would be measured from the values of factor loading that had been resulted from the factor analysis. An observed variable might be considered valid for the construct measurement (the latent variable) if the factor loading values had been above 0.3 [36]. Then, the construct reliability would be calculated from the number of standard square of loading factor from each indicator and the number of error variances ($\sum e_j$). The formula for measuring the values of construct reliability coefficient in a simple manner, as having been proposed by Setyo Hari Wijanto [37], would be as follows:

$$\text{Construct Reliability} = \frac{(\sum \text{std. loading})^2}{(\sum \text{std. loading})^2 + \sum e_j}$$

III. RESULTS AND DISCUSSIONS

The focus of the development measurement model of historical consciousness was to test the validity and reliability constructs of the instrument (the set of historical consciousness test) that had been designed and to test the goodness of fit the measurement model that had been designed. Then, the goal of the development was to attain empirical evidence regarding the factors and indicators that existed in the measurement model of historical consciousness.

The constructs of historical consciousness measurement instrument that had been hypothesis consisted of 4 factors and 19 indicators and the number of the test items was 90 items. From the analyzing process by means of confirmatory factory analysis the researcher found that out of 90 items that had been analyzed there were 28 items that had been invalid and insignificant (loading factor < 0.3 and t-value < 1.96). Therefore, the invalid test items would be eliminated from the analysis before the researcher re-analysis through model modification in accordance with the suggestions existed in the Lisrel software, so that the researcher would attain the fit model.

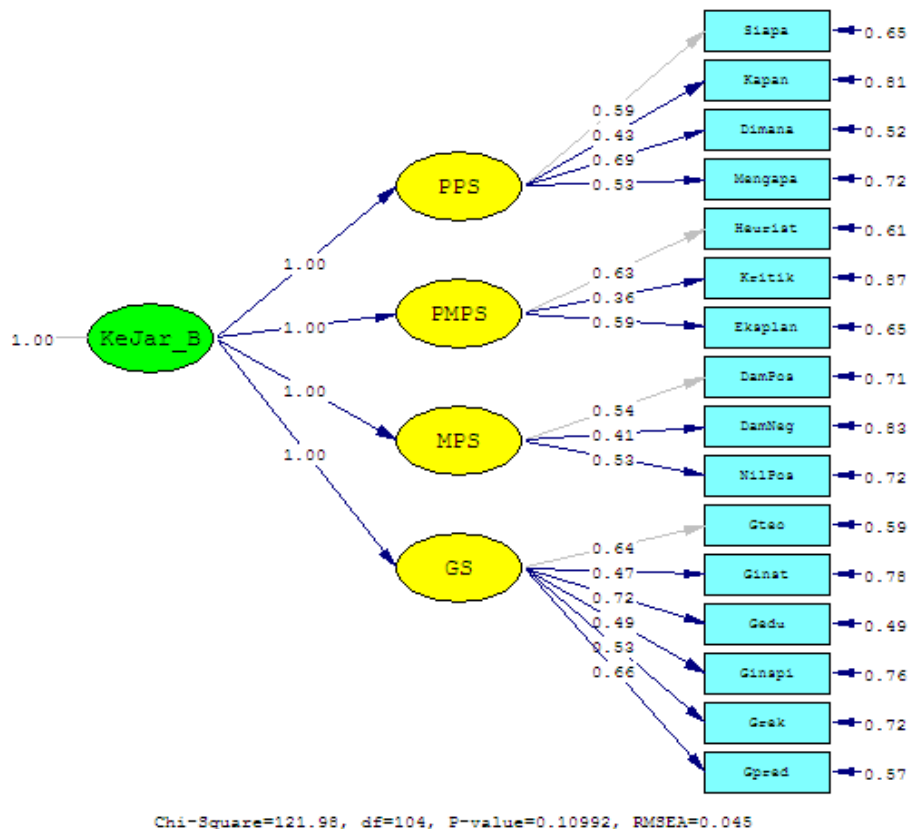


Figure 3. The Historical Consciousness Measurement Model Resulted from the Empirical Testing (Standardized)

The testing goodness of fit model prioritized the criteria that commonly used by viewing the Chi-Square value (χ^2 , the smaller the Chi-Square value was the better the result would be) and the probability value (p -value) ≥ 0.05 , RMSEA ≤ 0.08 [38]. The results of second order confirmatory factor analysis (2nd order CFA) showed that the measurement model of historical consciousness had been supported by the empirical data according to the p -value criteria (≤ 0.08). These findings proved that the measurement model in the historical consciousness test instrument showed the goodness of fit model and the constructs of the already hypothesized model had been accepted. Thereby, the measurement model of historical consciousness resulted from the theoretical review was supported by the empirical data (Figure 3)

An important aspect that should be given attention from the measurement model would be related to the validity and reliability of the constructs of measurement instrument. The validity and reliability of measurement model of historical consciousness were shown by the size of factor loading values from each indicators in the four latent variables (constructs) of historical consciousness. The results of 2nd order CFA as having been displayed in the Lisrel output (Figure 3) showed that the Standardized Factor Loading (SLF) value of the latent variable indicators of measurement model of historical consciousness had met the requirements, that is > 0.3 , and had been significant, as having been seen from the t -value that > 1.96 with the significance level 95%.

Table 4. The Construct Validity and Reliability of Test Instrument for the Historical Consciousness

Factor and Indicator		2 nd Order CFA		Construct Validity	Construct Reliability	
		SLF*	t-value		CR	Category
PPS	Siapa	0,59	**	Good	0,7	Good
	Kapan	0,43	3,48	Good		
	Dimana	0,69	5,02	Good		
	Mengapa	0,53	4,10	Good		
PMPS	Heuristik	0,63	**	Good	0,7	Good
	Kritik	0,36	2,99	Good		
	Eksplanasi	0,59	4,68	Good		
MPS	DamPos	0,54	**	Good	0,6	Acceptable
	DamNeg	0,41	3,16	Good		
	NilPos	0,53	3,86	Good		
GS	Gteo	0,64	**	Good	0,8	Good
	Ginst	0,47	3,89	Good		
	Gedu	0,72	5,54	Good		
	Ginspi	0,49	4,05	Good		
	GreK	0,53	4,29	Good		
	Gpred	0,66	5,16	Good		

Table 4 showed the size of factor loading values from each indicator that had passed the testing process of measurement model of historical consciousness. In overall, the factor loading values of all indicators ranged from 0.36 to 0.72. The lowest factor loading value was showed by the "criticism" indicator of the PMPS latent variable (0.36), while the highest factor loading value was found in the "Gedu" (GS) indicator from the GS latent variable (0.72). The t -value of all indicators ranged from 2.99 to 5.54 (>1.96). Therefore, it can be said that measurement instrument of historical consciousness in the form of associative multiple choice test items had good construct validity and had valid measurement model for measuring the historical consciousness. Then, it would be discussed the aspect of reliability.

The reliability of measurement model of historical consciousness was shown by the coefficient of Composite Reliability (CR). The composite reliability was known as multidimensional reliability because the measured constructs were multidimensional and were based on the confirmatory factor analysis. The coefficient of composite reliability explained the size of indicator proportion in explaining the measured constructs [39]. Coefficient of composite reliability had high accuracy in the multi-dimensional models [40].

The estimation on the coefficient of composite reliability for the constructs of historical consciousness measurement model in Table 32 ranged from 0.6 to 0.8. These coefficients were acceptable as long as the validity indicators of the model constructs were good [38]. The coefficient of CR for the dimension of PPS and of PMPS was equal to 0.7, the coefficient of CR for the dimension of MPS was equal to 0.6 and the coefficient of CR for the dimension of GS was

0.8. Therefore, it can be said that the historical consciousness measurement instrument might provide reliable (trustworthy) results.

IV. CONCLUSIONS

The results of the study show that the measurement model of historical consciousness that has been developed has been valid and reliable as well as fit to the empirical data. The constructs of measurement model of historical consciousness consist of four dimensions i.e. the knowledge of historical events, the understanding of historical research method, the meaning of historical event and the usefulness of history. The validity of measurement model is shown by the validity of constructs of test instrument through the factor loading values of all indicators in the measurement model of historical consciousness. The reliability of measurement model of historical consciousness is shown by the coefficient of composite reliability (CR) that is overall classified as good category. The empirical testing of fit model shows that the model has been fit, the χ^2 (Chi Square) = 121.98, the p -value = 0.11 and the RMSEA = 0.043. Thereby, it can be concluded that the measurement model of historical consciousness that has been developed has been supported by the empirical data.

REFERENCES

- [1] Rosenlun, D. (2011). *Subject construction, assessment and alignment in history education*. Paper presented in 39th Congres NPFP/ NERA di Universitas Malmö.
- [2] Kartodirdjo. (1993). *Pembangunan bangsa*. Yogyakarta: Aditya Media.
- [3] Thorp, R. (2014). *Historical consciousness, historical media, and history education*. Historiska medier. ISBN 978-91-7601-077-8 [versi elektronik].
- [4] Korber, A. (2015). *Historical consciousness, historical competencies – and beyond? Some conceptual development within German history didactics*. Paper has been presented at a conference in Oslo.
- [5] Bahri A. (2009). The Turkish prospective history teacher' understanding of analogy in history education. *International Journal of Historical Learning Teaching and Research*, 8, 1, 6-18.
- [6] Kementerian Pendidikan dan Kebudayaan. (2012). *Pemikiran tentang Pembinaan Kesadaran Sejarah*.
- [7] Mardapi, D. (2007). *Teknik penyusunan ins-trumen tes dan nontes*. Yogyakarta: Mitra Cendikia.
- [8] Greenberg, D. (1991). *Metahistory of everyday: historical consciousness in lived existence (set in late eighteenth century Britain)*. Tesis master, tidak diterbitkan, university of British Columbia. Canada.
- [9] Kartodirdjo. (1986). *Ungkapan-ungkapan filsafat sejarah Barat dan Timur: penjelasan berdasarkan kesadaran sejarah*. Jakarta: Gramedia.
- [10] Latief, J. A. (2006). *Manusia, filsafat dan sejarah*. Jakarta: Bumi Aksara.
- [11] Kreuzer, M. (2010). Historical knowledge and quantitative analysis: the case of the origins of proportional representation. *American Political Science Review*, 104 (2) 369-392.
- [12] Pompa, L. (1990). *Human nature and historical knowledge: Hume, Hegel and Vico*. Cambridge: Cambridge University Press.
- [13] Denison, B.J. (2011). *History, time, meaning and memory: idea for sociology of religious*. Leiden: Koninklijke Brill N Leiden: Koninklijke Brill NV.
- [14] Tucker, A. (Eds). (2009). *A companion to the philosophy of history and historiography*. Chichester West Sussex: Blackwell Publishing Ltd.
- [15] Pownall, F. (2007). *Lessons from the past: the moral use of history in fourth-century prose*. Michigan: University of Michigan Press.
- [16] Topolski, J. (1976). *Methodology of history*. (Terjemahan Olgierd Wojtasiewicz). Holland. Boston: D. Reidel Publishing Company. (buku asli diterbitkan tahun 1973).
- [17] Elliott, J. (2003). The limits of historical knowledge, *Eurepoen Review*, vol. II, pp. 21-13.
- [18] Grant, S.G. (2003). *History lessons: teaching, learning, and testing in US high school classrooms*. New Jersey: Lawrence Erlbaum Associate, Inc., Publisher.
- [19] Tosh, J. (1984). *The pursuit of history: aim, method, and new directions in the study of modern history (1st ed.)*. New York: Longman Inc.
- [20] Grant, S.G. (2003). *History lessons: teaching, learning, and testing in US high school classrooms*. New Jersey: Lawrence Erlbaum Associate, Inc., Publisher.
- [21] Kitson, A., Husbands, C., & Steward, S. (2011). *Teaching and learning history 11-18: understanding the past*. New York: Open University Press McGraw-Hill Education.
- [22] Paska, L. M. (2010). *Does film affect learning engagement?: historical inquiry and the document-based question in middle school social studies classroom*. Disertasi Doktor, Albany University, New York.
- [23] Tosh, J. (2002). *The pursuit of history: aim, method, and new directions in the study of modern history (3rd ed.)*. London: Longman Pearson Education.
- [24] Kuntowijoyo. (2013). *Pengantar ilmu sejarah*. Yogyakarta: Tiara Wacana.
- [25] Gottschalk, L. (1956). *Understanding history: a premier of historical method*. New York: Alfred A. Knopf, Inc.
- [26] Suhartono. (2010). *Teori dan metodologi sejarah*. Yogyakarta: Graha Ilmu.
- [27] Barash, J. A. (2003). *Martin Hiedegger and problem of historical meaning*. New York: Fordham University Press.

-
- [28] Russel, W.B., & Pellegrino, A. (2008). Constructing meaning from historical content: a research study. *Journal of Social Studies Research*, 32 (2) 3-15.
 - [29] Mazabow, G. (2003). *The development of historical consciousness in the teaching of history in South Africa School*. Doctor Dissertation, University of South Africa.
 - [30] Sjamsuddin, H. (2012). *Metodologi sejarah*. Yogyakarta: Ombak.
 - [31] Allen, M. J., & Yen, W.M. (1979). *Introduction to measurement theory*. California: Brooks/ Cole Publishing Company.
 - [32] Khine (Ed.). (2013). *Contemporary approaches to research in learning innovations: application of structural equation mode-ling in educational research and practice*. Rotterdam: Sense Publisher.
 - [33] Hendryadi. (2014). *Structural equation model-ing dengan Lisrel 8.80: pedoman untuk pemula*.Yogyakarta: Kaukaba.
 - [34] Kusnendi. (2008.) *model-model persamaan struktural*. Bandung: Alfabeta.
 - [35] Plomp, Tj & van de Wolde, J. 1982. The general model for systematical problem solving. From Tjeerd Plomp (Eds.). *Design of educational and training* (in Dutch). Utrecht (the Netherlands): Lemma. Netherland. Faculty of Educational Science and Technology, University of Twente. Enschede the Netherlands.
 - [36] Schumacker, R.E., & Lomax, R.G. (2010). *A Beginner's guide to structural equation modeling*. (3rded.). New York: Routledge.
 - [37] Wijanto, S. H. (2008). *Structural equation modeling dengan lisrel 8.8: konsep & tutorial*. Yogyakarta: Graha Ilmu.
 - [38] Hair, J. F., Black, W.C., Babin. B.J., & Anderson, R. E. (2007). *Multivariate data analysis*. (7th Ed). Prentice Hall. [versi elektronik].
 - [39] Margono, D. (2013). Aplikasi analisis faktor konformatori untuk menentukan reliabilitas multidimensi. *Statistika*, 13 (1) 17-24.
 - [40] Widhiarso, W., & Mardapi, D. (2010). Komparasi Ketepatan Estimasi Koefisien Reliabilitas Teori Skor Murni Klasik. *Jurnal Penelitian dan Evaluasi Pendidikan*, 14(1). Retrieved from <http://journal.uny.ac.id/index.php/jpep/article/view/1973>

ANBUSO: PRACTICAL SOFTWARE TO PERFORM ITEM ANALYSIS

Ali Muhson¹, Barkah Lestari², Supriyanto², Kiromim Baroroh²

¹Pendidikan Ekonomi FE Universitas Negeri Yogyakarta

²Pendidikan Ekonomi FE Universitas Negeri Yogyakarta

³Pendidikan Ekonomi FE Universitas Negeri Yogyakarta

⁴Pendidikan Ekonomi FE Universitas Negeri Yogyakarta
alimuchson@yahoo.com

Abstract—This study develops item-analysis software, and verifies its feasibility. This study uses a model of Research and Development (R & D). The procedure includes designing and developing a product, validating, and testing the product. The data were collected through documentations, questionnaires, and interviews. This study successfully developed item analysis software, namely AnBuso. It was practical and applicable for teachers to analyze test items.

Keywords: feasibility, item-analysis software, AnBuso

I. INTRODUCTION

Learning processes determine the quality of education. This implies that teachers have a very central role in improving the quality of education. They have crucial roles not only in designing and implementing but also in assessing the learning processes and their outcomes. Improving the quality of learning and assessment system is one of the efforts to improve the quality of education.

Learning and assessments can be said to be 'two sides of the same coin' [1]. All forms of assessment are essential components of teaching [2] and they have strong influences on the quality of learning [3]. Assessments involve not only finding out what students have learned but also improving the processes of learning and teaching [4]. This implies that assessments have essential roles in improving not only the learning processes but also the learning outcomes. Good assessments will encourage teachers to determine appropriate instructional strategies and motivate students to study harder.

An assessment is a process of gathering evidence of student learning to inform instructional decisions [5]. It means that teachers have to gather accurate evidence of the achievements of their students as the quality and impacts of their instructional decisions depend on it. The development of assessment instruments should therefore measure all aspects of learning objectives. Good students should attain better scores than the poor ones.

The implementation of the new curriculum, known as the 2013 curriculum, in the secondary education levels in Indonesia is facing many problems [6], for example the difficulty of changing teachers' mindset [7, 8]. A study reported that 87 percent of Indonesian teachers had difficulties in understanding the implementation of authentic assessments [9] as the assessment system includes too many aspects that must be measured [10]. It means that their ability to perform good assessments has to be improved.

An assessment is a systematic process of providing credible evidence of resources, implementation, and learning outcomes with the purpose of improving and developing effective instruction, programs, and services in higher education [11]. Thus, it is a process of gathering information about the students and in turn the information is employed to make instructional decisions and improve learning processes.

An assessment should be able to assess students' progress. Such information can be used as a basis for identifying the students' status in the group and determining the next steps of learning. Therefore, the assessment has to be able to accurately assess their competence. Assessments help teachers clarify the goals and the purposes of their teaching, create learning experiences in real-life contexts, and provide many different ways for students to demonstrate their abilities and skills [12]. Properly used assessment procedures can contribute directly to the

improvement of learning, such as (1) clarifying the nature of the intended learning outcomes, (2) providing short-term goals to work toward, (3) providing feedback concerning learning progress, (4) providing information for overcoming learning difficulties and for selecting future learning experiences, and (5) identifying the next instructional goal [13]. The procedures involve interrelated tasks in that one step determines the next.

The purposes of assessments include establishing a classroom that supports learning, planning and conducting instruction, placing students, providing feedback, diagnosing student problems and disabilities, and summarizing and grading academic learning and progress [2]. Thus, assessments are used to monitor the learning progress, provide information as a basis for giving feedback, and improve learning.

To provide accurate information, assessments have to be based on several criteria, such as valid, reliable, competence focused, comprehensive, objective, and educative [14]. The three concepts—validity, reliability, and objectivity—have to do with the quality of the information obtained from tests or other assessment instruments or methods [15, 16]. Thus, the validity and reliability are important parts of assessments to obtain appropriate and accurate information.

Item analyses play a somewhat more important role in the construct and predictive validation [17]. It means that the item analysis is an important process, which determines the validity of items and therefore it has an important role in identifying good items. The assessment results also provide accurate information about students' abilities therefore they can help teachers to identify which materials are difficult. The item analysis also provides information about the materials that have not been mastered by the students so that teachers can plan remedial activities. Based on the reasons it is necessary to develop a practical software that can help teachers analyze items practically.

There are several different types of item analysis software available in the market, but they have not been used optimally yet by Indonesian teachers. This is because most of the software use foreign languages which are difficult for them to understand. Some software are also less practical and applicable because they produce output with different formats. Therefore, it is necessary to develop more practical and applicable software that can cater for their needs. This present study aims to develop item analysis software that could be used by Indonesian teachers to prepare learning administration reports. The software is named AnBuso.

II. METHODS

This study uses a model of Research and Development (R & D). The procedure includes designing and developing a product, validating, and trying the product. A preliminary study was carried out to determine the ability and willingness of teachers to use item analysis and identify their real needs. The next step was designing a model of application software and its guidebook, which were validated by experts and revising them in order to obtain validated products.

The product validation was conducted using the Delphi method. This validation processes also involved some experts in education, measurements and computer programming. The testing of the trial products were conducted twice: limited and extensive testing. The test was to obtain comprehensive information and input from users. The results show that all the instruments, the software and its guidebook are feasible to use although there are some suggestions for revision. The results of this test were also used to improve the quality of the products so that they are ready to be disseminated widely.

The study involved 65 respondents comprising teachers and school supervisors from five districts in Yogyakarta (DIY). Purposive sampling was carried out to select the respondents with computer literacy, especially the ability to operate Microsoft Excel program.

Observation, documentation, questionnaires, and interviews were employed to elicit the research data. Observation was conducted to obtain data or information about the teachers' ability to use the application program that has been developed. The data of instruments used to measure students' abilities, such as formative and summative tests were obtained by documentation techniques. The elicited documents took the forms of final examination tests, daily tests, remedial programs, and the results of the previous item analysis used by the teachers.

Table 1. Software Feasibility Categories

No	Skor	Kategori
1	More than $M + 1,8 \text{ SD}$	Very feasible
2	$M + 0,6 \text{ SD}$ to $M + 1,8 \text{ SD}$	Feasible
3	$M - 0,6 \text{ SD}$ to $M + 0,6 \text{ SD}$	Fair
4	$M - 1,8 \text{ SD}$ to $M - 0,6 \text{ SD}$	Less feasible
5	Less than $M - 1,8 \text{ SD}$	Not feasible

The questionnaire elicited feedback or suggestion from teachers, supervisors, education department officers, and education experts. The questionnaire was also developed to measure the feasibility of the software concerning its visual display, contents, usefulness, practicality, and convenience. Interviews were conducted to teachers, supervisors, education department officers, and experts to obtain information regarding the strength and drawbacks of the AnBuso software.

The data were analysed descriptively and supported by tables and graphs. Five scales of categories were employed to measure the feasibility of the software (see Table 1).

III. FINDING AND DISCUSSION

The study involved 65 participants comprising teachers, supervisors and education actors. Most of them (72%) were teachers of high schools (SMA, SMK, and MA). Many of them (68%) were teachers of public schools and 57% were government employees. Respondents came from various fields of study, including Economics, Accounting, Mathematics, Social Studies, English, Indonesian, Information Technology, Chemistry, Physics, Biology, Geography, Arabic, and so on.

The willingness of the teachers to analyze test items was low (57%). Only 11% of them regularly conducted test item analysis and 12% of them never did it. It would be important to assess why the teachers had a low willingness to analyze test items.

In general, the teachers would perform item analysis if the supervisors ordered them to do so. It means that their awareness to perform item analysis was very low. This is because they were unable to use the available software. Of course this contributed to the poor quality of test items they have developed.

This study successfully developed AnBuso software and its manual book. AnBuso was developed on Microsoft Excel program. It has been improved based on the feedback from the respondents of the present research. The AnBuso was accompanied with a manual book in Indonesian language that will help teachers take advantage of the software. It provides step-by-step instructions which are very practical for the teachers to operate the software. The book contains an introduction, contents, input and data report.

The software has undertaken some important modifications in order to be compatible with the assessments system of the 2013 curriculum. Some revisions were conducted to improve its visual display, **Input01** spreadsheet, **Participant Reports** spreadsheet, **Participant Remedial** spreadsheet, and formula changes. The modifications on the visual displays had to be done because it had too many color variations, which have made the application less attractive. The adjustments were done not only on the input spreadsheet, but also on the report spreadsheet (see Figure 1). They were made based on the respondents' feedback.

AnBuso 5.0		
© 2011-2014 by Ali Muhson		
IDENTITAS		
(Hanya diperkenankan mengisi data atau menghapus tetapi tidak boleh memindah isi data atau menggunakan fasilitas Cut Paste)		
Data Umum	Kolom Pengisian	VALIDASI
1 Satuan Pendidikan	SMA Negeri 80 Yogyakarta	OK
2 Mata Pelajaran	Ekonomi	OK
3 Kelas/Program	XI IPS	OK
4 Nama Tes	Sumatif	OK
5 SKKD		Belum Diisi
6 Nama Guru	Ali Muhson	OK
7 NIP	196811121999031003	OK
8 Semester	Gasal	OK
9 Tahun Pelajaran	2012/2013	OK
10 Tanggal Tes	12 Agustus 2012	OK
11 Tanggal Diperiksa	7 September 2012	OK
12 Nama Kepala Sekolah	Mustofa	OK
13 NIP Kepala Sekolah	1972031111984031003	OK

AnBuso 5.1 For Teacher		
© 2011-2014 by Ali Muhson		
(Hanya diperkenankan mengisi data atau menghapus tetapi tidak boleh memindah isi data atau menggunakan fasilitas Cut Paste)		
Keterangan	Kolom Pengisian	VALIDASI
1 Nama Kepala Sekolah	Mustofa, M.Sc.	OK
2 NIP Kepala Sekolah	1972031111984031003	OK
3 Tanggal Laporan	Yogyakarta	OK
4 Tanggal Laporan	20 Agustus 2014	OK
5 Sikala Penilaian (4, 10 atau 100)	4	OK
6 Nilai KKM	72.66	OK
Jumlah dan Bobot Soal		
7 Jumlah soal pilihan ganda (Max 50)	20	OK
8 Jumlah soal essay (Max 10)	5	OK
9 Bobot soal pilihan ganda	60%	OK
10 Bobot soal essay	40%	OK

Previous visual display
New visual display
Figure 1. The spreadsheet of Input01 before and after modifications

AnBuso 5.1 For Teacher		
© 2011-2014 by Ali Muhson		
(Hanya diperkenankan mengisi data atau menghapus tetapi tidak boleh memindah isi data atau menggunakan fasilitas Cut Paste)		
Keferangan	Kolom Pengisian	VALIDASI
at Laporan	Yogyakarta	OK
al Laporan	20 Agustus 2014	OK
Skala Penilaian (4, 10 atau 100)	4	OK
Nilai KKM	2.66	OK
Jumlah dan Bobot Soal		
Jumlah soal pilihan ganda (Max 50)	20	OK
Jumlah soal essay (Max 10)	5	OK
Bobot soal pilihan ganda	60%	OK
Bobot soal essay	40%	OK
Data Soal Pilihan Ganda		
Jumlah Alternatif Jawaban (Max 5)	4	OK
<input type="button" value="Input 01"/> <input type="button" value="Input 02"/> <input type="button" value="Lap Peserta"/> <input type="button" value="Lap Butir"/> <input type="button" value="Pola Jawaban Butir"/> <input type="button" value="Lap Essay"/> <input type="button" value="Materi Remedial"/> <input type="button" value="Pese"/>		

Figure 2. Grade Revision

AnBuso has been adjusted to assessment system of the 2013 curriculum of the secondary education levels in Indonesia so that it accommodates Indonesian teachers' needs. Some amendments were also made on the assessment scales. In the previous model, the grading scales were 1-10 and 1-100. To cater for the assessments system of the 2013 curriculum, the new model provides a 1-4 grade scale (see Figure 2).

Basically the purpose of the item analysis is to determine the quality of the test items and obtain information about the assessment results so that teachers can make a follow-up plan such as remedial and enrichment teaching. The software also provides information about the students who are included in the remedial and enrichment group.

Figure 3 shows the general overview of the AnBuso system. Basically the software is divided into three components: input, process and output. Data input consists of **Input 01** which takes general data, and **Input 02** which contains the identity of students and their answers. The data entered into the **Input 01** spreadsheet include the School Name, Subject, Class/Program, Name of test, SK/KD, Name of the Teacher, Teacher's NIP, Semester, Academic Year, Test Date, Inspected Date, name of the school Principal, Principal's NIP, Reporting Place, Reporting Date, Grades (10 or 100), the score of Minimum competence criteria (KKM), Objective Test Data, and Essay Test Data. The data entered into the **Input 02** spreadsheet include Students Name, Gender (M = Male, P = Female), Multiple Choice Answer sheet, and Essay Score. Whenever the data are entered in the input 01 and input 02 spreadsheets, the program will automatically generate a complete report consisting of:

1. **Student Report Sheet** which displays list of test scores. The list contains the name of student and gender, number of items answered correctly and incorrectly, objective test score, essay test score, final score, and description (passed and failed). The other information in this sheet is the total number of the students, the number (percentage) of students who passed the tests, the minimum scores, maximum scores, average scores, and the standard deviation.

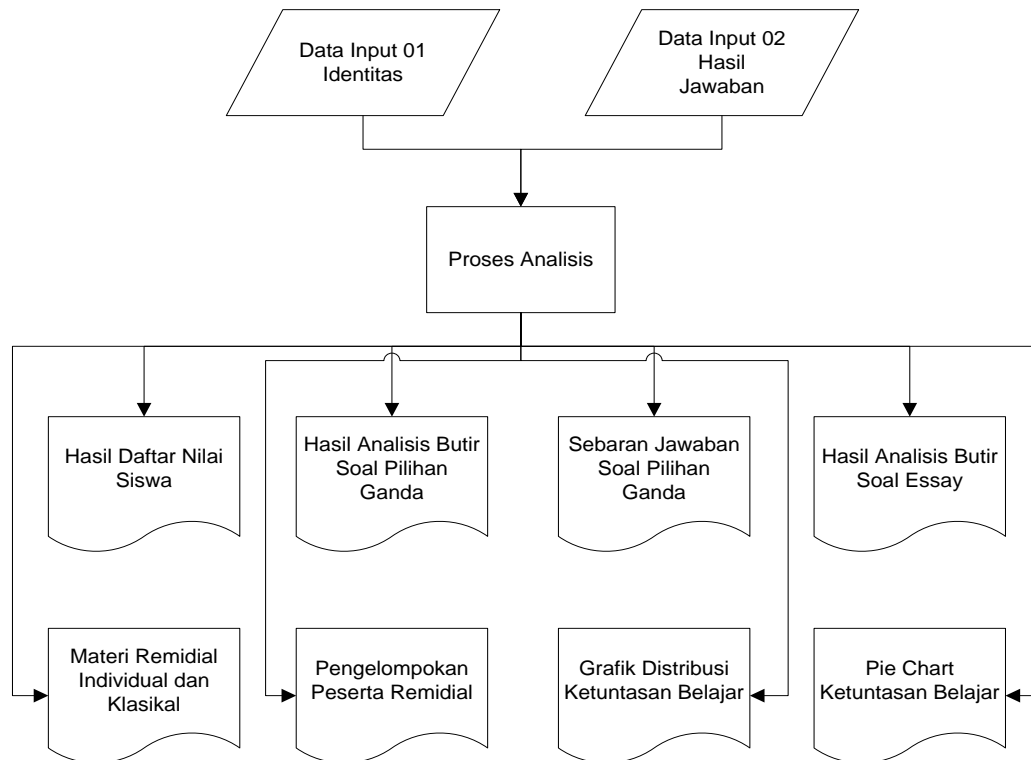


Figure 3. Flowchart Diagram Revision

2. **Item Report Sheet** which shows the result of Multiple Choice Analysis. It contains Discriminant Index, levels of difficulty, Malfunction Distractors, and Conclusion.
3. **Response Pattern Report Sheet** which contains Distribution of the answers of Multiple Choice tests.
4. **Essay Report Sheet** that contains Discriminant Index, levels of difficulty, and Conclusion.
5. **Remedial Material Reports Sheet**, which contains materials for Individual and Classical Remedial tests.
6. **Remedial Group Reports Sheet**, which contains the grouping of participants who take remedial tests.
7. **Graphics Sheet**, which contains charts of minimum competence criteria (KKM) score distributions, and the proportions of the learning mastery.

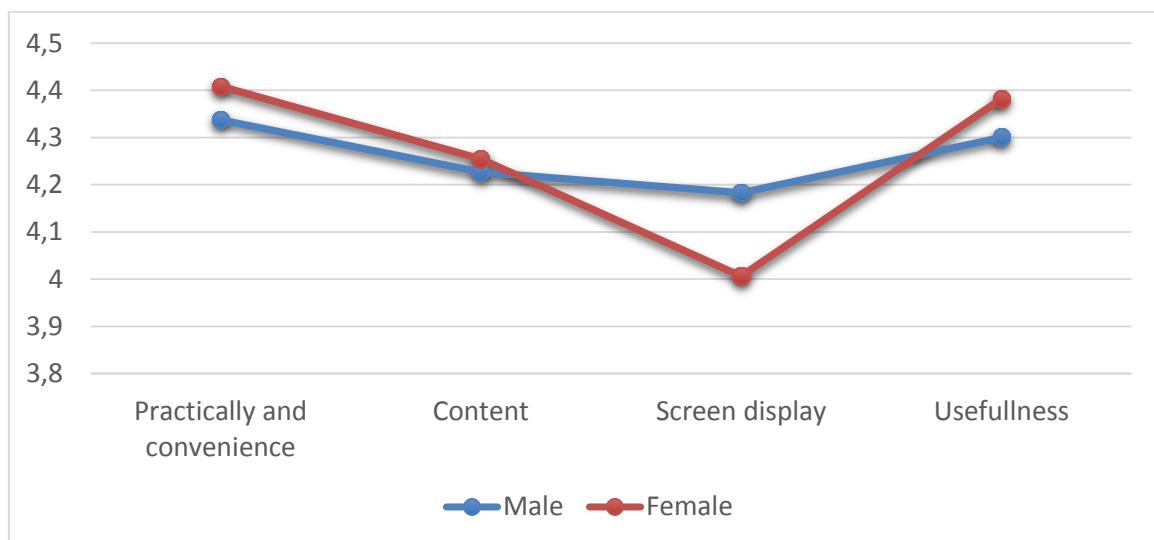


Figure 4. The results of the feasibility test

The results of the feasibility test showed that AnBuso software, developed in this study, was considered very feasible by the respondents. Of the total respondents, 51% said that it is feasible, 46% stated it is very feasible and the rest stated it is quite feasible. This shows that the software has a high usefulness in helping teachers perform item analysis.

The aspect that gets the highest score is its practicality or convenience, and usefulness. The lowest rated aspect is the visual graphic display (see Figure 4). As the respondents reported, the AnBuso is very practical to assist them analysing test items and it really provides them with complete information. The software is able to analyze not only multiple choice test items but also essay tests. The results of the test-item analysis are presented in the forms of report, therefore it is very easy for teachers to interpret them.

IV. CONCLUSION

This study successfully developed AnBuso, which can be used by Indonesian teachers to analyze items practically. This software is developed on Microsoft Excel program which contains data input spreadsheet, data processing spreadsheet, analysis report sheet in tables and chart. AnBuso is considered as a very feasible software by the teachers regarding its practicality and convenience, usefulness, and contents, while its screen display is considered only feasible. As AnBuso is very feasible and very useful, it should be disseminated more widely to teachers. This software needs to be upgraded periodically in order to fulfill teachers' future need.

V. REFERENCES

- J. T. Guliker, T. J. Bastianes and P. A. Kirschner, "A Five-Dimensional Framework for Authentic Assessment," 1] *Educational Technology Research and Development*, vol. 52, no. 3, pp. 67-86, 2004.
- M. K. Russell and P. W. Airasian, *Classroom assessment: concepts and applications* (7th edition), New York: McGraw-Hill, 2012.
- J. E. Raymond, C. S. Homer, R. Smith and J. E. Gray, "Learning through authentic assessment: An evaluation of a 3] new development in the undergraduate midwifery curriculum," *Nurse Education in Practice*, vol. 13, no. 5, pp. 471-476, 2012.
- T. H. Bers, "The role of institutional assessment in assessing student learning outcomes," *New Directions for 4] Higher Education*, vol. 141, pp. 31-39, 2008.
- R. J. Stiggins and J. Chappuis, *An introduction to student involved assessment for learning*. Sixth edition, Boston: Pearson assessment training institute, 2012.
- Republika, "Implementasi kurikulum 2013 masih dibayangi banyak masalah," 2014. [Online]. Available: 6] www.republika.co.id. [Accessed 18 November 2014].
- Metronews, "Ini delapan masalah dalam implementasi kurikulum 2013," 2014. [Online]. Available: 7] [News.metronews.com](http://news.metronews.com). [Accessed 19 October 2014].
- Tempo, "Problematika implementasi kurikulum 2013," 2013. [Online]. Available: www.tempo.co. [Accessed 10 July 8] 2013].
- Susilowati, "Kurikulum 2013, 87 persen guru kesulitan cara penilaian," 2013. [Online]. Available: <http://unnes.ac.id>. 9] [Accessed 18 November 2014].
- Tempo, "Kurikulum 2013, Apa Saja Kendalanya?," 2014. [Online]. Available: www.tempo.co. [Accessed 16 August 10] 2014].
- T. W. Banta and C. A. Palomba, *Assessment essentials: Planning, implementing, and improving assessment in 11] higher education*, San Fransisco: Jossey-Bass, 2014.
- L. Darling-Hammond, *Next generation assessment: Moving beyond the bubble test to support 21st century 12] learning*, San Fransisco: Jossey-Bass, 2014.
- M. D. Miller, R. L. Linn and N. E. Gronlund, *Measurement and assessment in teaching* (tenth edition), New Jersey: 13] Pearson Education Inc., 2009.
- Puskur, *Model Penilaian Kelas Kurikulum Berbasis Kompetensi*, Jakarta: Balitbang Depdiknas, 2008. 14]
- L. W. Anderson, *Classroom assessment: enhancing the quality of teacher decision making*, New Jersey: Lawrence 15] Erlbaum Associates, Inc., 2003.
- T. Kubiszyn and G. D. Borich, *Educational testing and measurement: classroom application and practice*. 10th 16] edition, New Jersey: John Wiley & Sons, Inc., 2013.
- J. C. Nunnally and I. H. Bernstein, *Psychometric theory* (third edition), New York: McGraw-Hill, Inc., 1994. 17]

ESTIMATING OF STUDENTS CAPABILITY GROWTH IN VERTICAL EQUATING WITH RASCH MODEL TEST

Anak Agung Purwa Antara

IKIP Saraswati Pahlawan Street no 2 Tabanan 82113, Phone.(0361) 811267

Email: purwa.antara@gmail.com

ABSTRACT- This study aims to find: (1) the development of students' ability of class III, IV, V, and VI Elementary School in math based on tests used; (2) accuracy mean and sigma method on vertical equating used in this study; and (3) the quality of the test in this study developed by rasch model analysis. The research data were students scorer on summative tests elementary school 2015 in Tabanan Bali collected with stratified random sampling. Design of equating using anchor test design, and the test instrument using the rasch model analysis. The results showed that (1) the ability of students in math has increased by an average θ class III, IV, V, and VI, respectively for -0.1152, -0.0657, 0.0096 and 0.2834; (2) the value Root Mean Square Difference (RMSD) methods mean and sigma 0.2347; and (3) the quality of the four test developed quite well with indicators of the average value of the logit item 0.00 showing the instrument is measuring, the value Infit MNSQ and Outfit MNSQ approaching 1.00, the value Infit ZSTD and Outfit ZSTD approaching 0.00, separation successively 5.40, 5.43, 5.52, 5.57, that means four tests have good slope, and the fourth test of reliability quite well with the value 0.9700, 0.9300, 0.9500, and 0.9400.

Keywords: *Ability Development of Students, Rasch Model Tests.*

I. BACKGROUND

Graduates of educational programs in accordance with the Law on the National Education System 20 of 2003 should include three competencies, namely the attitude, knowledge and abilities, so it can be beneficial for himself, for others, and for the universe and all its contents and its civilization. Helpful in this regard formulated by indicators faithful, righteous, noble, healthy, knowledgeable, skilled, creative, independent, and become citizens of a democratic and accountable. The implication of the statement is any educational program organized by an educational institution must be able to produce quality resources, so as to compete in facing the challenges of life both now and in the future.

In order for these objectives to be realized, any educational program organized by an educational institution should always be monitored and evaluated for improvement, both in planning and in implementation so that the educational goals can be achieved.

The curriculum is basically designed lesson plans based on the objectives of education. As a good learning planning system, the curriculum must include four things: (1) the end result to be achieved learning program learners (output), which is defined as the competence of graduates; (2) the content of the material to be taught and learned by the student (input / content standards), in forming a desired competence of graduates; (3) the implementation of learning (process, including teaching methodology as part of the standard process) so that third desirable competencies formed on students; and (4) the conformity assessment process, and achievement of learning goals as early as possible to ensure that input, process, and outcomes as planned.

In this regard Mardapi (2008, p. 5) said that, improving the quality of graduate or graduate competence of an educational institution can be done through improving the quality of learning programs and improving the quality of teacher assessment done in the classroom. This implies that teachers as a component in direct contact with the student should be able to increase the competence - competence of the students, either through the implementation of learning programs, as well as through the assessment. Assessment models used by teachers, should be

able to monitor or provide information about the learning process has been done, the development of the ability of students after attending the learning, and the learning outcomes that have been achieved by students.

Good judgment requires the correct data that needs to be supported by a quality measurement process (Mardapi, 2008, p. 5). Each institution can implement a measurement program at grade level and school-level test either formative or summative tests. All these measures are expected to inform the development of improved learning outcomes neighbor either on grade level, school level and national level so that it can be done in-depth studies in order to improve the quality of graduates.

Implementation of quality measurement program would encourage all components involved in the learning process, especially teachers and students to make efforts in achieving quality learning outcomes in accordance with the expected goals. Teachers required to increase their knowledge and broaden mastery of the subject matter so as to make an assessment and develop measurement tools (tests) are qualified to be used to measure achievement of learning programs that have been conducted. Through quality tests expected to obtain accurate information about the outcomes of learning and development of students' abilities in teaching.

Information about the development of students' abilities in a learning program organized by an educational institution needs to be known in full. One way to determine this is through vertical equating. Equating vertical can be used by teachers to unveil the development of students' abilities, even if the student has a level different capabilities and are at different grade levels as long as the tests used to measure the trait of the same (Croctor & Algina, 1986, p. 456; Hambleton & Swaminathan, 1985, p. 197).

The use of vertical equating at different grade levels in a subject will help teachers to get information about the development of students' abilities on those lessons to the whole class rank. The ability of students increases will impact on increasing the quality of learning outcomes or graduates. It is therefore important to conduct vertical equating for an educational institution because through vertical equating can be obtained information about the development of students' abilities and various other trends. Such information would be a reference in the restoration effort towards enhancing the quality of learning outcomes or graduates in these institutions.

The problem to be answered in this study were (1) how the development of students' ability of class III, IV, V, and VI Elementary School in math based on tests used?; (2) how the accuracy of the mean and sigma method on vertical equating Rasch model tests used in this study?; (3) how the quality test with Rasch models that developed in this study?

Theoretically, this research is expected to contribute to the development of measurement in education, among others: (1) provide information on the vertical equating with test instruments Rasch models is more developed for SMP and SMA / SMK; (2) provides information on the ability of students at different grade levels so that it can be used as a reference for the evaluation of learning in the classroom; and (3) provide information to develop research on vertical equivalency tests for fields studi other IPA.

In practical results of this study can be utilized in: (1) search growth of student achievement in a course of learning periodically according to class rank; (2) perform a comparison item difficulty, especially mathematics according to the assessment period and class rank, so it is possible the shift of learning materials according to class rank; and (3) do the majors through the selection of individual students' ability levels.

II. RESEARCH METHODS

This study begins with the development of instruments (tests) mathematics achievement multiple choice for Class III, IV, V, and VI Elementary School tested in the 2nd half (summative test) which is based on the subject of numbers, geometry and measurement. The fourth test package contains a grain of anchor as much as 20%. Preparation of the lattice and-writing done by a team consisting of two senior teachers of elementary school mathematics courses. Test the validity of the content using techniques Gregory (2000, p. 123). While testing the suitability of measurement hypothetical model to empirical data (the construct validity) using the Rasch model analysis. The program used is winsteps (Linacre, 1994, p. 328) with indicator infit-outfit meansquare, outfit and infit zstandard, reliability item, item separation, and unidimensionality. Criteria Outfit Mean Square is $0.5 < \text{MNSQ} < 1.5$; Z-Standard Outfit value received is $-2.0 < \text{ZSTD}$

<2.0 and dimensionality accepted criteria is the raw measure of variance explained by $\geq 20\%$ and unexplained variance <10%.

Collecting data using stratified random sampling study involving a sample of 800 elementary school students of class III, IV, V, and VI in Tabanan Bali. Application of random done at the school level, whereas the determination of strata school to consider the location of the school and school categories.

Equating method using the mean and sigma and calibration methods together (concurrent calibration) with the design of test items common anchor or non-equivalent groups designs. The accuracy of the method mean and sigma tested with Root Mean Square Difference (RMSD) the ability of the student (θ) before and after equating (Kim & Cohen, 2002, p. 31), information about the development of students' abilities in math can be obtained from the equalizing score and scale capabilities of students in fourth grade were involved.

III. RESULTS

Analysis test developed include test content validity and construct validity. Content validity of test class III, IV, V, and VI, respectively for 0778, 0711, 0756, and 0733 are all above 0700 (the criteria used CV > 0.700). Thus the four tests that are prepared in compliance with the specified criteria.

The construct validity of the experimental data using winsteps program are presented in Table 1. Table 1 shows that the quality of the four test developed quite well. This can be seen from the indicators mean value logit item around 0.00 that indicates the instrument is measuring. Infit MNSQ and Outfit MNSQ approaching 1.00, and the Infit ZSTD and Outfit ZSTD approaching midnight showing of good quality. Value of separation successively 5.40, 5.43, 5.52, 5.57, which means all four tests have good slope, and the value of Cronbach Alpha respectively 0.97, 0.93, 0.95, and 0.94 showing the fourth test has good reliability.

Table 1. Summary of Running Winsteps

Component		Test class III	Test class IV	Test class V	Test class VI
<i>Mean of logit item</i>		0.00	0.00	0.00	0.00
<i>Meansquare (MNSQ)</i>	<i>Infit</i>	1.00	1.00	1.00	1.00
	<i>Outfit</i>	1.01	1.00	0.99	1.00
<i>Zstandard (ZSTD)</i>	<i>Infit</i>	0.00	0.00	0.10	-0.10
	<i>Outfit</i>	-0.10	-0.10	0.00	0.00
<i>Item separation</i>		5.40	5.43	5.52	5.57
<i>Dimensionality</i>	<i>Raw variance explained by measure $\geq 20\%$</i>	22.9%	22.2%	20.3%	24.1%
	<i>Unexplned variance in 1st contrast</i>	7.7%	7.1%	9.7%	6.0%
	<i>Unexplned variance in 2nd contrast</i>	6.5%	6.0%	6.7%	5.5%
	<i>Unexplned variance in 3rd contrast</i>	5.9%	5.6%	6.0%	5.2%
	<i>Unexplned variance in 4th contrast</i>	5.7%	4.9%	4.8%	5.0%
	<i>Unexplned variance in 5th contrast</i>	5.5%	4.6%	4.2%	4.8%
<i>Alpha Cronbach</i>		0.97	0.93	0.95	0.94

The constant equating α and β of the mean & sigma method calculated the mean and standard deviation of item difficulty anchor. Results equating capability parameter θ are presented in Table 2. Accuracy equating is calculated based on the value of RMSD ability before and after equating. The value of the mean and sigma method RMSD obtained 0.2347.

Table 2. Mean and Standard Deviation θ Equating Results

Equating	Mean θ	Standard Deviation
Theta Class III	-7.8E-07	1.002
Conversion θ (III to IV)	-0.191	1.011
Conversion θ (III to V)	-0.351	1.372
Conversion θ (III to VI)	-0.484	0.843
Theta Class IV	-1.2E-06	1.002
Conversion θ (IV to V)	-0.026	0.798
Conversion θ (IV to VI)	-0.257	0.708
Theta Class V	-1.2E-17	1.002
Conversion θ (V to VI)	-0.259	1.039

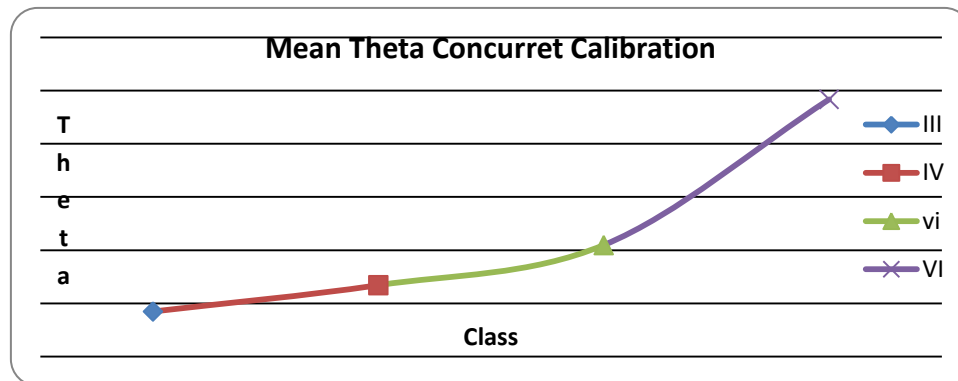


Figure 1. Graphics Improved ability of students

Average Traffic students appear to be declining if converted on a scale on it. It can be interpreted that the average student in class VI Traffic higher than average ability students of class V. The mean ability of students in grade V is higher than the average ability of fourth graders and fourth graders Traffic mean higher than average ability students of class III. This is reinforced by the average ratio of students Traffic concurrent calibration results that show the magnitude of average ability students of class III, IV, V, and VI, respectively for -0.1152, -0.0657, 0.0096 and 0.2834. Graph increase students' abilities shown in the figure below.

IV. CONCLUSIONS AND RECOMMENDATIONS

Results equalization separately using the method and the results mean and sigma equalization simultaneously (concurrent) showed that the average ability of students of class III, IV, V, and VI Elementary School in Tabanan Bali has increased along with the increasing class rank. The highest increase occurred in the sixth grade. Mean & sigma method used in this study is quite accurate with RMSD value of 0.2347.

Quality test developed quite well with indicators of the average value of the item logit approaching midnight. Value Infit MNSQ and Outfit MNSQ approaching 1:00, value Infit ZSTD and Outfit ZSTD approached midnight, the value of separation in a row 5:40, 5:43, 5:52, 5:57 and Cronbach Alpha value respectively 0.97, 0.93, 0.95, and 0.94.

Advised the teachers to constantly monitor the development of the ability of students in each subject. Teachers need to be trained to analyze tests that can provide accurate information to the learning outcomes that have been done.

REFERENCES

- [1] Croker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehard and Winston Inc.
- [2] Gregory, R.J. (2007). *Psychological testing. history, principles, and applications* (4th ed). Boston: Pearson Education, Inc.
- [3] Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- [4] Kim, S.H., & Cohen, A.S. (2002). A comparison of linking and concurrent calibration under the grade response model. *Applied Psychological Measurement*. 26, 25-61.
- [5] Linarc, J.M. (1994). Sample size and item calibration stability. *Rasch measurement transactions*. 88, 355-383
- [6] Mardapi, D. (2008). *Teknik penyusunan tes dan nontes*. Yogyakarta: Mitra Cendikia

DIAGNOSTIC TEST CHARACTERISTICS OF LEARNING DIFFICULTIES IN MATHEMATICS FOR SCIENCE CLASS 12th GRADER

Apri Triana⁽¹⁾, Heri Retnawati⁽²⁾

Departement of Educational Research and Evaluation, PPs UNY
email: apri_triana@yahoo.com⁽¹⁾, retnawati_heriuny@yahoo.co.id⁽²⁾

Abstract- This development research aims to determine characteristic of diagnostic test for science class 12th grader math learning difficulties. The diagnostic test is developed work to find out where the students difficulty for science class 12th grader according indicators on mathematics domains are developed in adaptation of National Examinations Competency Standard For Senior High School. Math domain test are Algebra, Geometry, Trigonometry and Calculus. The model used in the development of diagnostic test difficulty learning mathematics is DINA model (deterministic input, noisy " and " gate). On the development of test with DINA model, items test accompanied by the determination of attribute. Attribute is ability to be possessed by students completed the items correctly. Attribute that make each item test consists of content attributes and process attributes. Results of this research are: (1) An expert judgment which analysed using formula Aiken, result 26 good items and 4 items that should be replaced; (2) Analysis of trials out produced 26 good items, and 4 items were not good; (3) Analysis of field test diagnostic produce 29 good items, and one item is not good because difficulty level of 2.01; (4) Coefficient reliability test is 0,97. That means diagnostic test has highly reliable to measure math skills for science class 12th grader on domain Algebra, Trigonometry, Geometry and Calculus in accordance with the National Examinations Competency Standard For Senior High School; (5) there are 29 good items test, 14 items have wrong answers more than 50%.

Keywords: *diagnostic test, DINA, item characteristic, and reliability*

I. INTRODUCTION

Based on the results of the UN year 2013/2014 as mathematics class XII Science had a mean value of 6,04. Mean value was lower than the average score of English language that is equal to 6,47. While, average value of Bahasa Indonesia reached 7,16 and 6,43 to mean of Physics. The mean low scores on mathematics, can be lack of students overcome with domain math.

Command of the mathematics must important, because the students which has good mathematical ability must understand math concepts in a hierarchically structured, logical, and systematic. The concepts of mathematics that requires students to be more creative in develop ability to mathematical thinking. Because of demands mathematical thinking, many students difficult in math problems solving. In addition, many students have difficulty in reading and understanding the symbols in mathematics. Solve of Maths which abstract and written in symbolic can made difficulties for students.

The difficulties of mathematics learning can be diagnosed if developed diagnostic good test. The diagnostic test useful test to determine learning difficulties by students, belong the errors of concept understanding. Performed test can be used, when the information was obtained that most students fail following the process of learning (Djemari Mardapi, 2012: 112).

Based on interview with one of the mathematics teacher for favorite high school in Metro. Remedial have just provide retest, but given't repetition domain math which difficult. The impact of remedial retest, will make students emphasis concepts of math are less and lead difficulties for learning in domain mathematics. In addition, tested often mathematics learning are formative test and summative tests. The function of formative test and summative test certainly different diagnostic test.

Research diagnostic difficulties of learning have done before there are Fauzan (2010), Suwanto (2011), and M. Duskri (2014). While, research on the DINA model has also been used by Kusaeri (2012) in the development of diagnostic tests of Algebra, and also used for analysis National Examinations Elementary School year 2012/2013 by Widi Wulansari (2014). From some

research about diagnostics, researchers have not found development of diagnostic test for National Examinations Competency Standard For Senior High School. For that, researchers are interested in developing diagnostic test for science class 12th grader math learning difficulties using DINA Model.

Based on some issues and difficulties students in learning math, need for diagnosis in subject matter of mathematics. Command of mathematics students to be improved, especially the preconditions domain, an interrelated domain math. Development of diagnostic tests don't often in mathematics learning difficulties, in this research will discussed about Diagnostic Test Characteristics Of Learning Difficulties In Mathematics For Science Class 12th Grader.

II. CONTAINING THE BACKGROUND

There are still difficulties of domain math for science class 12th grader and often diagnostic test aren't developed or used in learning of mathematics.

III. RESEARCH PROBLEM

Problems in this research development how was characteristics of diagnostic test mathematics developed for science class 12th grader in Metro according with adaptation of National Examinations Competency Standard For Senior High School. This Research using analysis with Rasch models, because there is only attention difficulty of item. Findings difficulties of mathematics learning for Science Class 12th Grader, will presenting of students' mistakes percentage in according procedure solve by indicators and domain tested.

IV. OBJECTIVES

This Research objective was determine characteristics of diagnostic test which developed in adaptation of National Examinations Competency Standard For Science Class 12th Grader on domain Algebra, Geometry, Trigonometry and Calculus. In addition, this research aims to determine of students' mathematics learning difficulties percentage for science class 12th grader in according with domain tested.

V. SIGNIFICANCE OF RESEARCH

Kusaeri (2012) has research with title *Dina In Development Model Using Diagnostic Tests For Detecting Wrong Conception*. This research aims to 1) find out a way to develop the diagnostic test using the DINA model, so that it can give information about misconceptions in algebra, 2) Identify the characteristics of the good diagnostic test developed by using the DINA model. This research and development is an empirical research with the descriptive explanatory used to describe the development stages of the test. The subjects of this research were year VIII students of SMPN 1 Yogyakarta, SMPN 1 Sanden Bantul, and SMPN 1 Panjatan, Kulon Progo.

The data were analyzed by using the CDM, Mplus, and R software. The results of the study are as follows 1) The stages of the test development in this research were: identifying basic competence and formulating indicators, constructing the learning continuum, constructing the domain hierarchy, formulating the attributes, constructing the problems, conducting validation by expert judgment, and administering an empirical test. Through those seven stages, 37 items of the diagnostic test were developed, 2) Of the 37 items, 15 items must be eliminated/discarded from the test. The items were eliminated because their quality was low and they did not meet the requirements of the model fit test because their discrimination indexes were less than 0.2.

VI. RESEARCH METHODS/MODEL

This research is development research, which was to development diagnostic test difficulty learning of mathematics. In process, carried out empirical research. Empirical research used descriptive exploratory. Descriptive research can using describe stages of development diagnostic tests. Exploratory used to finding information of characteristic item are difficulty items along with reliability instrument.

Used of model in this research is *deterministic-input, noisy "and" gate* (DINA) model. DINA Model is psychometric models used to evaluate of strengths and weaknesses for students (Jimmy de la Torre, 2009: 1). In DINA model, probability that examinee i correctly answers item j ($Y_{ij} = 1$) is expressed in the DINA model as follows (Koken Ozaki, 2015):

$$P(Y_{ij} = 1 | \alpha_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}$$

Where α_i be the knowledge state of examinee i and;

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$$

Here, q_{jk} is the element of the Q-matrix for item j . If attribute k is needed to answer item j , then $q_{jk} = 1$, and otherwise $q_{jk} = 0$. Furthermore, s_j be the *slip* parameter of item j , which is the probability that an examinee who has all of the knowledge needed to correctly answer the item fails to answer it correctly. While, g_j be the *guessing* parameter for item j , which is the probability that an examinee who does not have all of the knowledge needed to correctly answer the item nevertheless does correctly answer it. In the DINA model, the slip and guessing parameters of item j are defined as $s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$ and $g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$, respectively (Jimmy de la Torre, 2009)

DINA Model is one of cognitive diagnosis models (CDM) which used to evaluate the strengths and weaknesses of students (Jimmy de la Torre, 2009). With an approach that has been done by Tatsuka, that underlying the student's cognitive ability is the ability of the process and understanding necessary to answer the test items on the Matrix-Q, where the columns of the matrix expressed cognitive ability and lines represent grain or vice versa (Tatsuoka, 2009: 53), Knowledge and cognitive skills and understanding of the ability of the process according to Tatsuka called attributes (2009).

To the end, step for development diagnostic after made items test, next step is determine attributes of underlying these items. The attributes that underlying grain of diagnostic tests only can used in this research. After attributes are identified, atribut will be structured in hierarchy. Define Hierarchy is psychological order among the attributes required to problems solve test. The order attributes derived from empirical considerations (Leighton, Gierl, & Hunka, 2002).

The attribute hierarchy method for cognitive assessment (AHM; Leighton et al., 2004) is a psychometric method for classifying examinees' test item responses into a set of structured attribute patterns associated with different components specified in a cognitive model of task performance. This method illustrates how cognitive performance can be evaluated using an information processing approach because the AHM requires a cognitive model of structured attributes to evaluate examinee performance. These structured attributes, called the attribute hierarchy, define the psychological ordering among the attributes required to solve a test problem and, thus, serve as the cognitive model of task performance. The ordering of the attributes may be derived from cognitive considerations about the attribute structure (Gierl, Leighton, and Hunka: 2007).

VII. CONTAINING THE APPROACH USED

Development Research of in this paper using quantitative approach. Quantitative deskriptif for this research used explore steps of development instrument diagnostic test and characteristic test. This research using *deterministic-input, noisy "and" gate* (DINA) model.

VIII. POPULATION AND SAMPLE

The population in this research were all students of science class 12th grader in Metro. While the research samples are 323 students of science class 12th grader of State Senior High School in Metro. In the out trials of instruments, samples of out trials were 97 students of science class 12th grader of Senior High School Number 9 Yogyakarta.

IX. INSTRUMENT

This research using tes of instrument. Form of test is multiple choice which have 30 item test. For each items there are five alternative answers. Objective tests used to aim diagnosis students' difficulties learning, then each alternative answers given predetermined possible difficulties that maybe experienced by students. Meanwhile, for content validity, examinee of each item made composed sheet of domain, basic competencies, indicators and questions.

X. SOURCE OF DATA

Trial out of instruments test conducted in Senior High School Number 9 Yogyakarta. Subject trials out of instrument test are 97 students of science class 12th grader. From 97 students on trials out test, 21 students working two packages of test (62 student working of packet 1 and 56 student working of packet 2). After trials out, analysis and repairs instrument test, next step is field test. Samples of field test were 323 student of science class 12th grader. Trials out Instrument conducted in Senior High School Number 9 Yogyakarta, while the field test was conducting six State of Senior High School in Metro.

XI. TECHNIQUE OF DATA ANALYSIS

This Research used analysis with formula Aiken to determine size of index validity and coefficient Cronbach Alpha to determine reliability instrument test. Determine of characteristics instruments diagnostic test which developed, it used by QUEST program with approach Rasch model. For knowing percentage of students' difficulties mathematics learning can used calculation manual by excel.

Content validity test used to estimate feasibility or relevance of content test through rational analysis by competent panel or *expert judgment*. Analysis used to determine of coefficient content validity is Aiken formula. The following formula of Aiken (Saifuddin Azwar, 2012: 113);

$$V = \frac{\sum s}{n(c-1)} \quad (1)$$

With V is index validity of Aiken; s is difference between criterion- i with the lowest rating. In this research have four rating criteria (1, 2, 3, 4) with lowest rating criteria is 1. Number of rating is c , while n is the number of expert. Calculation coefficient of reliability instrument used Cronbach Alpha formula (Djemari, 2012: 72):

$$r_i = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_{x_{ig}}^2}{\sigma_y^2} \right) \quad (2)$$

Based on equation (2), is the coefficient reliability instrument with number of k item. Variance score item stated of $\sigma_{x_{ig}}^2$ and σ_y^2 as variance of total score. To know characteristics of item developed diagnostic tests, carried out calculations approach with Rasch model. Shape Rasch model equation in measurement is (Heri retnawati, 20 114: 14-15):

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} \quad (3)$$

From equation (3), it is probability test participants have ability θ to selected randomly item can answer item- i correct. The parameters in this model is that an index difficulty of item- i . While e is natural numbers which values approaching 2,718. Parameter level of difficulty item both from -2.0 to +2.0. For a value near -2.0 then categorized as very easy, while a value close of +2.0 then categorized as very difficult (Hambleton, 1991: 13).

XII. RESEARCH FINDINGS AND DISCUSSIONS

The domain for selected items in diagnostic test developed by absorption of domain results from National Examinations of Mathematics For Senior High School in 2014/2015 are still low. Based on domain has been identified difficulty of students, then selected indicators that have percentage of the low absorption with National Examinations of Mathematics. Results from summary is based on absorption of 14 indicators obtained National Examinations of Mathematics For Senior High School still has low absorbency percentage.

Based on 14 indicators specified percentage of low absorptive capacity, subsequently reduced to 30 items form of multiple choice. Each item in diagnostic test was made attributes that underlie each item. Attribute is the ability must be possessed by students to answer correctly each item. The attributes of each item then arranged in a matrix hirearchy (Attribute hierarchy method).

After items of diagnostic test are assembled into device test, then every items by expert validated. As for which is validated by expert is compatibility between the indicator with items. Results of the analyzed expert validation with Aiken formula. From analysis by Aiken, obtained four items that should be revised or replaced due to lack of appropriate indicators or expert

opinion. After the items are replaced or revised, device diagnostic test and then trials out instrument test.

Trials out instruments conducted for science class 12th grader in Senior High School Number 9 Yogyakarta. Based on test results of instruments test, there are four items not good criteria of difficulty item. Fourth the item was later replaced with different item for same indicators. After instrument test repair, next step in this research is field test research.

Developed test in field test research was performed involving six of State for Senior High School in Metro. Based on analysis conducted research, acquired characteristics of the instrument test are; (1) from 30 items tested, produced an item that doesn't good criteria because it has difficulty item level of 2:01; (2) Coefficient reliability test is 0.97, that means has highly reliable diagnostic test to measure math skills for science class 12th graders domain on Algebra, Trigonometry, Geometry and Calculus in accordance with the National Examinations Competency Standard For Senior High School.

The result test results showed of high percentage of students' difficulties learning. From 29 items were good, there are 14 items that false percentage of more than 50% students and 14 items still have false percentage of students more than 35%. Results of classification difficulty based on the domain tested, percentage of students' Learning Difficulties based on false answer items test are still high. Percentage mean of false on domain Calculus of 54.49%, percentage mean of false on domain Algebra amounted to 50.06%, 48.92% percentage mean of false on domain Trigonometry and Geometry have percentage mean false amounted to 46.75%.

XIII. CONCLUSIONS AND SUGGESTIONS

Based on the discussion of results obtained some conclusions. First, measures used development of diagnostic tests in this research to determine purpose/objectives test, determine domain test, determine indicators and made items from these indicators, define attributes of each item, carry out expert validation, repair item of test, trials out, revise, field test, analysis and assemble final product of diagnostic tests. Second, the final product in this research is multiple choice test with 29 items as five alternative answers. Third, average percentage false of students math still high for each domain, which amounted to 54.49% false in Calculus, false of Algebra amounted to 50.06%, false amounted to 48.92% in Trigonometry and false in Geometry is 46.75%.

Results of this reseach, expected there are repairs and improvements for future research. The next reseach similary have given more Items for test and give more variance question, so more can diagnose difficultly of mathematics learning. Indicators of National Examinations Competency Standard For Senior High School can be lowered to item in diagnostic testing, in order to give more information about the Learning Difficulties of mathematics. For next research, is expected to more subject, it could be more representative. Also tried to develop a test for the same model in this research.

REFERENCES

- [1] Djemari Mardapi. (2012). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika
- [2] Fauzan. (2010). *The Development of a Diagnostic Test for Mathematics Learning Difficulties in Senior High Schools*. Dissertation. Yogyakarta: Graduate School, Yogyakarta State University.
- [3] Gierl, M. J. (2007). *Making Diagnostic Inferences about Cognitive Attributes Using the Rule-Space Model and Attribute Hierarchy Method*. Journal of Educational Measurement, Vol. 44, No. 4, IRT-Based Cognitive Diagnostic Models and Related Methods (Winter, 2007), pp. 325-340.
- [4] Hambleton, Ronald K., Swaminathan H., Rogers Jane H. (1991). *Fundamental of Item Respon Theory*. United States of America: Sage Publications
- [5] Heri Retnawati. (2014). *Teori Respon Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika
- [6] Jimmy de la Torre. (2009). *DINA Model and Parameter Estimation: A Didactic*. Journal of Educational and Behavioral Statistics March 2009, Vol. 34, No. 1, pp. 115-130. DOI: 10.3102/1076998607309474
- [7] Koken Ozaki. (2015). *DINA Models for Multiple-Choice Butirs With Few Parameters: Considering Incorrect Answers*. Applied Psychological Measurement XX(X) 1-17. DOI: 10.1177/0146621615574693
- [8] Kusaeri. (2012). *Dina In Development Model Using Diagnostic Tests For Detecting Wrong Conception*. Educational Research and Evaluation Journal, Year 16, Number 1, 2012. Yogyakarta State University.
- [9] Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2002). *The Attribute Hierarchy Model for Cognitive Assessment*. Retrieved from <https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwiykuOvo7rMAhVSB04KHeJu>

- AxkQFgg8MAM&url=http%3A%2F%2Fwww.crame.ualberta.ca%2Ffiles%2Fahm_ncme2002final.pdf&usg=AFQjCN
EtWs2xHlatWphvUJERTm3dH9jYPA&sig2=IZrfjQfc5LTyUZ1y3YDlig&bvm=bv.121070826,d.c2E.
- [10] M. Duskri. (2014). *Developing Diagnostic Test Of Mathematics Learning Difficulties In Elementary Schools*. Educational Research and Evaluation Journal, Year 18, Number 1. Yogyakarta State University.
 - [11] Saifuddin Azwar. (2015). *Reliabilitas dan Validitas: Edisi 4*. Yogyakarta: Pustaka Pelajar.
 - [12] Suwato. (2011). *Developing a Diagnostic Test to reveal Student's Difficulties in their Understanding of the Cell Reproduction*. Dissertation. Yogyakarta: Graduate School, Yogyakarta State University.
 - [13] Tatsuoaka, K. K. (2009). *Cognitive Assessment: An Introduction to the Rule Space Method*. New York: Taylor & Francis Group, LCC.
 - [14] Widi Wulansari. (2014). *An Analysis for Finding Information on the Erroneous Concepts of Students in Problem Solving Based on Results of National Examination (NE) Elementary School Mathematics Year 2012/2013 in Kediri*. Thesis. Yogyakarta: Graduate School, Yogyakarta State University.

ASSESSING SCIENCE PROCESS SKILLS USING TESTLET INSTRUMENT

Ari Syahidul Shidiq^{*1}, Sri Yamtinah² dan Mohammad Masykuri²,

¹ Student of Science Education Master Program, Sebelas Maret University, Surakarta Indonesia

² Lecturer of Science Education Master Program, Sebelas Maret University Surakarta, Indonesia
*email: arisyahidul@yahoo.co.id

Abstract-The aim of this research was to develop an instrument to assess student science process skills. A paper and pencil objective test that has been developed in form of Testlet and for specific chemistry content. The science process skills indicators has been discussed and defined with expert judgment using focus group discussion (FGD) technique. The indicators of science process skills are conceptual knowledge, observation, controlling variable, interpreting data and making conclusions. The test consisted of 30 Testlet multiple choice items and given to 75 student of 11th grade from different senior high school. This research used 3 representative schools, that represented high, middle and low level in Surakarta. The result from the research are the total test reliability was measured at 0,843 using chronbach's alpha, the mean and the standard deviation for the test are 16,04 and 5,50, the average percentage correctly answered by student from high, middle and low level are 61,1%, 34,9%, and 24,9%.

Keywords: *Science process skills, Testlet, instrument, assessment*

I. INTRODUCTION

Nowadays, the aim of education and science education is to educate individuals in order to have skills that is needed in the future. Skills can help students to adapt with the environment. For instance, students can think flexible, creative, critical, tolerant and respect the others opinion, and they can use the skills of science process to solve the problems [1].

Nowadays the most discussed issue of science education is assessment of science process skills. Science process skills are special skills that simplify learning science, activate students, develop students' sense of responsibility in their own learning, increase the permanency of learning, and teach them the research methods. Besides that, science process skill is used to get information, think on problems and formulate the results. This skill is important for the students to learn how to apply concept, theory, and law that is accepted in learning science. Science process skill is appropriate for all knowledge. The skill has been the important part in science curriculum. It has also been the effective approach in teaching learning process. Therefore, the curriculum development that is oriented on science process skill needs the reliable and valid instrument development which can evaluate the increase of the skill [2,3,4]

Science process skills can be categorized into two, basis and integrated (high process skills). Base science process skills includes observation, classification, communication, measurement, prediction, and intervention. Meanwhile, High science process skills (integrated) includes manipulating, interpreting, defining operationally, formulating a model, conducting experiment, designing hypothesis and making conclusion [5,6]

Science process skills that does not include assessment will give effect on the useless learning. Therefore, it is needed to conduct assessment development that is appropriate with science process skills [7]. There are many researches on science process skills such as the research about application and assessment of science process skills conducted by S. U. Gezer, 2015; J. Ambros, L. Meiring and S. Blignaut, 2014; K. Kurea-In and O. Thongperm, 2013; C. Keil, J. Haney, and J. Zoffel, 2009; D. R. Lavoie, 1999; W. Foulds and J. Rowe, 1996; P. J. German, 1994; W. M. Roth and A. Roychoudhury, 1993. Research on the assessment of integrated science process skills conducted by K. G. Tobin and W. Capie, 1982; F. G. Dillashaw and J. R. Okey, 1980; R. H. Yeany., K. C. Yap., and M. J. Padilla, 1986; B. E. Weasley., G. H. Krockover., and A. Devito, 1985; J. C. Burns, 1985 and E. H. M. Sahali., And L. Halim, 2010 [8-20].

Science process skills is an inter-disciplined knowledge. All science topics are integrated to each other. Therefore, it is also needed an instrument that can measure science process skills and it can be applied to science material that is integrated and hierarchy. The alternative choice instrument of testlet can be applied in this case.

Testlet is a group or item group that is related with the certain topic that is developed into a unity. It consists of some steps that is determined before and can be followed by the participants. Testlet belongs to a kind of test that results more than one response. Furthermore, Testlet has response that relatively rises with a relevancy to knowledge that will be measured [21].

II. METHODOLOGY

The research began with Focus Group Discussion (FGD) to determine the Science Process Skills (SPS) indicators. Nine expert judges, consist of 5 senior teachers and 4 lecturers reviewed the initial indicators and test draft. The result of FGD indicators that are used in this experiment are conceptual knowledge, conducting observation, controlling variable, interpreting data and making conclusions. The number of items for each SPS indicators are listed in Table 1.

Table 1. Numbers of item for each SPS indicators

No	SPS indicators	Item Number	Number of Item
1	Conceptual knowledge	1, 4, 7, 10, 16, 22, 28,	7
2	Observation	8, 11, 14, 20, 29	5
3	Controlling variable	6, 12, 17, 23, 25, 26	6
4	Interpreting data	2, 5, 13, 19, 21	5
5	Making conclusions	3, 9, 15, 18, 24, 27, 30	7

The test consisted of 30 Testlet multiple choice items and given to 75 student of 11th grade from different senior high school. This research used 3 representative schools, it represented high, middle and low level in Surakarta. The test was conducted for 90 minutes. The data of this preliminary test was used to investigate the SPS of the students and to determine test reliability, item discrimination and item difficulty

III. RESULTS

The obtained data was used to determine the item analysis and descriptive statistic of the test. Reliability of the test was measured at 0,843. It means that the test has high degree of consistency and stability in measuring what it is intended to measure. Calculation of reliability used Alpha Crobach's coefficient. The skewness was measured at 0,346. It means that the data has normal distribution. The item analysis of item difficulty and item discrimination are listed in Tabel 2.

Table 2. Descriptive Statistic

No	Descriptive Statistic	Result
1	Number of Item	30
2	Number of Examinees	75
3	Mean	16,040
4	Variance	30,252
5	Standar Deviation	5,500
6	Skewness	0,346
7	Kurtosis	-0,758
8	Minimum	7,000
9	Maximum	29,000
10	Median	16,000
11	Alpha	0,843

The item analysis of item difficulty and item discrimination are listed in Tabel 3.

The item discrimination of number 1, 7, 11 and 29 indicate the low result of discrimination. It means that the item is not appropriate to differentiate between high achiever and low achievers student. Item number 11 and 29 shows the negative result. It means that group of students with high achievement have less correct answer than group of student with low achievement. However, 26 of 30 items are appropriate to the purpose of the test. The item difficulty indicates the range from 0,067 (the most difficulty item, that is item number 21) to 0,960 (the easiest item,

that is item number 7). Based on result of item discrimination and item difficulty of the item number 11 and 29, it cannot be used or must be dropped from the total test, and item number 1, 7, and 21 can be used but must be repaired in order to get better result.

Table 3. Item Analysis for Difficulty and Discrimination

Item Number	Item Discrimination	Item difficulty	Item Number	Item Discrimination	Item difficulty
1	0,036	0,613	16	0,588	0,560
2	0,392	0,800	17	0,314	0,840
3	0,288	0,760	18	0,387	0,653
4	0,145	0,920	19	0,359	0,107
5	0,598	0,227	20	0,585	0,533
6	0,663	0,200	21	0,251	0,067
7	0,001	0,960	22	0,159	0,813
8	0,566	0,680	23	0,741	0,333
9	0,587	0,587	24	0,568	0,307
10	0,582	0,587	25	0,667	0,440
11	-0,024	0,147	26	0,740	0,493
12	0,452	0,573	27	0,368	0,307
13	0,217	0,920	28	0,462	0,693
14	0,416	0,733	29	-0,065	0,493
15	0,642	0,307	30	0,562	0,387

The test that consists of 30 testlet multiple choice items is given to the different level of schools that is high, middle and low level. Each school took one class as a respondent. All of representative schools applied scientific approach. The result of the test can be compared to know the student science process skills from each level. Percentage of student that can master all of SPS indicators from high, middle and low level of school are 61,1%, 34,4% and 24,9%. Conceptual knowledge as a basic SPS indicates the best result of SPS that student can master, on the other hand the students cannot still show a good result in interpreting data and making conclusion. Moreover, controlling variable indicator can be only correctly answered by 1% student in low level schools. The average result of 4 SPS indicator (except conceptual knowledge) indicates that the students who answer correctly are less than 50 %. The result indicates that student cannot apply and process conceptual knowledge to different condition. The students only know the conceptual knowledge without understanding how to apply, process and link the knowledge to the new condition, make conclusion and solve the problems as aspected by science process skills. The results of student Science Process Skills are shown in FIG.1.

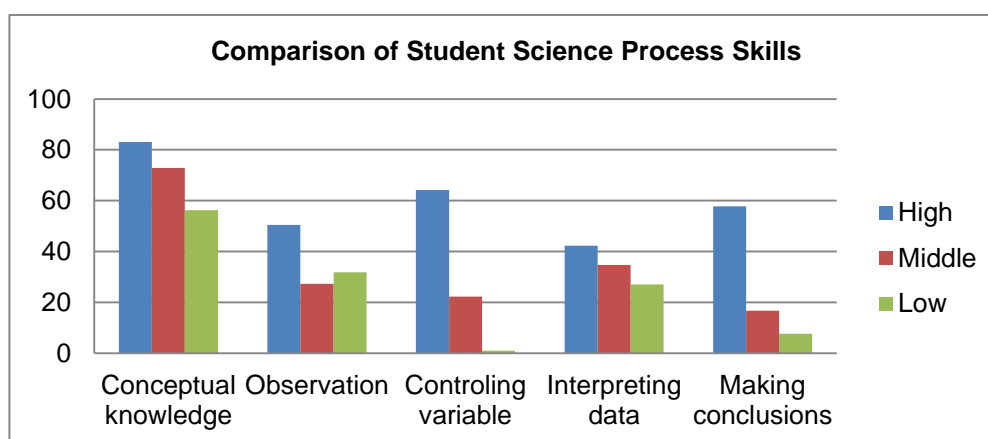


Figure 1. Comparison of Student Science Process Skills

IV. CONCLUSION

Science process skills are important for students to learn how to apply concept, theory, and law that they received while they are studying science. These skills need to be taught to students through scientific based learning. However, learning science process skills while excluding its assessment will result in a meaningless learning. Therefore, assessment is necessary

in learning science process skills. One of the ways to realize that is through this research. This research uses a testlet instrument that could be used as an alternative instrument for science process skills assessment.

The result of analysis toward testlet instrument shows a good result. However, it still needs to be improved and developed in order to give a better and more accurate result, so it can be used as a proper instrument to assess science process skills.

The result of science process skills analysis that used five indicators from this research indicates that students from 3 different school levels show a positive result in conceptual knowledge indicator but not in the other four indicators (observation, controlling variable, interpreting data and making conclusions). This is caused by students that have yet to comprehend how to apply, process, and associate conceptual knowledge for making conclusion and solving problem. Therefore it is necessary for students to be given an instrument that can encourage student's mind to think scientifically, so that students have integrated conceptual knowledge, observation, controlling variable, interpreting data, and making conclusions skills.

REFERENCES

- [1] H. Aktamis and N. Yenice. Determination of the science process skills and critical thinking skills levels. *Procedia, Social Behavioral Science*. 2. pp 3282-3288. 2010
- [2] M. Hodosyova, J. Utla, M. Vanyova, P. Vnukova and V. Lapitkova. The development of science process skills in Physics education. *Procedia, Social Behavioral Science*. 186. pp 982-989. 2015
- [3] S. Karamustafaoglu, Improving the Science Process Skills Ability of Science Student Teachers Using I diagram, *Eurasian J. Phys. Educ.* 3 (1): 26-38, 2011
- [4] E. H. M. Shahali and L. Halim. Development and Validation of Test of Integrated Science Process Skills, *Procedia, Social Behavioral Science*, 9, pp 142-146, 2010
- [5] W. Lati, S. Supasorn and V. Promarak. Enhancement of learning achievement and integrated science process skills using inquiry learning activities of chemical reaction. *Procedia, Social Behavioral Science*. 46. pp 4471-4475. 2012
- [6] A. Aydin. Representation of science process skills in the chemistry curricula for grade 10, 11 and 12/ Turkey. *International Journal of Education and Practice*, vol 1. No. 5. pp 51-56. 2013
- [7] W. Harlen. *Assessment & Inquiry - Based Science Education: Issues in Policy and Practice*. Trieste: The Global Network of Science Academies Science Education Programme. 2013
- [8] S. U. Gezer. A case study on preservice teacher laboratory usage self efficacy and scientific process skills. *Procedia, Social Behavioral Science*. 174. Pp 1158-1165. 2015
- [9] J. Ambross, L. Meiring and S. Blignaut, The implementation and development of science process skills in the natural science: A case study of teachers perception, *Africa Educational Review*, 11:3, pp 459-474. 2014
- [10] N. Kurea-In and O. Thongperm. Teaching of Science Skills in Thai Contexts: Status, Supports and Obstacles, *Procedia, Social and Behavioral Science*, 141. pp 1324-1329, 2014
- [11] C. Keil, J. Haney, and J. Zoffel. Improvements in Student achievement and science process skills using environmental health problem-based learning curricula. *Electronic Journal of Science Education*. Vol 13. No. 1. 2009
- [12] D. R. Lavoie. Effect of emphasizing hypothetic-predictive reasoning within the science learning cycle on high school student's process skills and conceptual understanding in biology. *Journal of Research in Science Teaching*. Vol 36. No. 10. pp 1127-1147. 1999
- [13] W. Foulds and J. Rowe. The enhancement of science process skills in primary teacher education student. *Australian Journal of Teacher Education*. Vol 21. No 1. 1996
- [14] P. J. Germann. Testing a model of science skills acquisition: An interaction with Parents education, preferred language, science attitude, cognitive development, academic ability, and biology knowledge. *Journal of Research in Science Teaching*. Vol 31, no 7. pp 749-783. 1994
- [15] W. M. Roth and A. Roychoudhury. Development of Science Process skills in authentic contexts. *Journal of Research in Science Teaching*. Vol 30. No 2. Pp127-152. 1993.
- [16] K. G. Tobin and W. Capie. Relationship between formal reasoning ability, locus of control, academic engagement and integrated process skills achievement. *Journal of Research in Science Teaching*. Vol 19. No.2. pp 113-121. 1982
- [17] F. G. Dillshaw and J. R. Okey. Test of integrated science process skills for secondary science student. *Science education journal*. Vol 65. No. 5. Pp 601-608. 1980
- [18] R. H. Yeany, K. C. Yap and M. J. Padilla. Analyzing hierarchical relationship among modes of cognitive reasoning and integrated science process skills. *Journal of Research in Science Teaching*. Vol 3. No 4. Pp 277-291. 1986
- [19] B. E. Wesley, G. H. Krockover, and A. Devito. The effect of computer assisted instruction and locus of control upon preservice elementary teachers acquisition of the integrated science process skills. *Journal of Research in Science Teaching*. Vol 22. No. 8 pp 687-697. 1985
- [20] J.C Burns, J. R. Okey, and K. C. Wise. Development of an integrated process skills test: TIPS II. *Journal of Research in Science and Teaching*. Vol 22. No. 2. Pp 169-177. 1985
- [21] D. Thissen and H. Wainer, H. *Test Scoring*. Mahwah: Lawrence Erlbaum, 2001

The Effect of Multiple Choice Scoring Methods And Risk Taking Attitude toward Chemistry Learning Outcomes

(An Experiment at SMA Negeri 13 Kota Bekasi, West Java)

Awaluddin Tjalla¹, Sari Fitriani²

¹Jakarta State University, Indonesia

²Jakarta State University, Indonesia

awaluddin.tjalla@yahoo.com, fysaree@gmail.com

Abstract— The objectives of this research is to investigate: (1) the differences chemistry learning outcomes between students who are treated with multiple choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, (2) the interaction effect between scoring methods on multiple choice tests and risk taking attitude toward chemistry learning outcomes, (3) the differences chemistry learning outcomes between students who are categorized as risk seeking and treated with multiple choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, and (4) the differences chemistry learning outcomes between students who are categorized as risk averse and treated with multiple choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method. The research was conducted in SMAN 13 Kota Bekasi. The samples of this research were 40 students of XI grade in Science Program Class during second semester of the academic year 2015/ 2016. The samples were selected by simple random sampling technique. The method of the research used an experimental method. While the design of the research used treatment by level 2x2. Data collection was conducted by measuring student risk taking attitude and student chemistry learning outcomes. The hypotheses of research were tested using two way analysis of variance (ANOVA). The result of the research has a positive implication in the effort of learning evaluation system enhancement in senior high school and also student chemistry learning outcomes enhancement.

Keywords: *Multiple Choice, Scoring Method, NRET, FS, Risk Taking Attitude*

I. INTRODUCTION

Chemistry is the foundation of the natural sciences such as, medical science, pharmacy, geology, engineering, and others. Therefore, chemistry is an important part of science that has a major role in the development of a nation. But until recently, studied chemistry in high school was considered difficult by students, it is caused by the characteristics of chemistry itself, those are: chemistry teaching materials that are abstract, chemistry is a simplification of the actual situation, chemistry teaching materials are sequential and growing rapidly, learning chemistry is not only to solve problems but students also must learn the descriptions of the facts of chemistry, the rules of chemistry, the terms of chemistry, and materials studied in chemistry are very much (Middlecamp & Kean, 1985: 9). The difficulty of studying chemistry impacts on students' learning outcomes in senior high school, which is still relatively low.

Based on observations that was conducted at SMAN 13 Kota Bekasi, showed that the results of studying chemistry in senior high school is still low in terms of achieving minimum criteria score (KKM) and the acquisition of the national examination score. Average acquisition of the national examination score of chemistry for the last three years at SMAN 13 Kota Bekasi is still relatively low, which is on the classification C and D.

Table 1. Average of the National Examination Score of Chemistry at SMAN 13 Kota Bekasi for the Last Three Years

Academic Year	2012/ 2013	2013/ 2014	2014/ 2015
Clasification	C	D	C
Average	6,45	5,25	6,77
Lowest	4,00	3,25	4,00
Highest	9,25	8,50	9,00
Standar Deviation	1,26	1,24	1,15

(Source: List Collective of National Exam Results (DKHUN) SMAN 13 Kota Bekasi)

Less satisfying results of studying chemistry in senior high school can be caused by several factors. Factors that influence learning outcomes can be either internal or external (Suryabrata, 2007: 233). One factor that is believed to be able to influence the results of studying chemistry is the selection of the test item format in evaluation program that is applied by teachers in the classroom, especially in the formative evaluation. Formative evaluation carried out not merely provide score as a result of the measurement process, but it gives the meaning of the score that achieved by students, and at the stage of reflection, teachers can motivate students for the further improvement of the learning process so that the achievement of competencies as defined in learning objectives can be seen in the results obtained by student.

Multiple choice test item format is the most common used format test in various educational evaluation program, it is because multiple choice can be scored easily and quickly, objectively, more easily analyzed, may include extensive material on a test, can measure the ability of a variety from the simplest to the most complex. In the formative evaluation program, a multiple choice test is generally scored with conventional scoring methods called number right scoring method (NR). In conventional scoring method (NR scoring method), each test score is the sum of the item scores for a given examinee, and the examinee is awarded one point for the correct item response and zero for any other response (Crocker & Algina, 1986: 399). The main concern in this scoring method is that students can answer correctly by guessing (Choppin, 1988: 384-386). Number of correct answers on a conventional scoring method can consist of two things; the number of questions where the students actually knew the answer, and the number of questions which the students guess the answer correctly. Sometimes, students only know the part of the answer or answered uncertainly, it is known as partial knowledge (Coombs, Milholland, Womer, 1956: 36). Therefore, when using this conventional scoring method, the teacher cannot distinguish whether the student answers correctly based on the knowledge and ability or based on guesses. Guesses that benefit will boost students' scores thus causing overestimate the ability of the student.

Using NR scoring method in multiple choice test, not only partial knowledge of students is not credited, but teachers also cannot diagnose students' misunderstanding and lack of understanding in order to provide informative feedback to facilitate students' continuous learning (Lau, et al, 2011: 99). Additionally Kulhavey, et al said that students spend less time learning for multiple choice tests relative to an essay tests (Roediger & Marsh, 2005: 1155). Nana Sudjana argues objective test has been considered as an easy test to answer and students lack of preparation for the exam at school (Sudjana, 2009: 43). It shows that the multiple choice test with conventional scoring methods are less stimulating students' motivation, so it would be very influential on student learning outcomes.

The weaknesses of multiple choice tests mentioned make chemistry teachers in senior high school tend to only use essay test format than multiple choice test format in formative evaluation. But then another problem arises, as the Indonesian Government Regulation number 74 of 2008 concerning teacher, explained that the workload of teachers must at least meet the 24-hour face to face and at most 40 hours of face to face in one week at one or more units of educational institutions that have a license from the government or local government. It means chemistry teachers must teach about 240 until 400 students every week (public schools filled of 40 students per class). Based on that regulation, when teachers always implement essay test format in formative evaluation, then teachers will be very busy correcting test answer sheets in every single office day, considering essay test takes longer and difficult for correcting than the multiple choice test. Furthermore, if the correction tests take so long, then the feedback is expected from the formative evaluation could become irrelevant for the students, because of the time between the tests given and feedback given is too far away.

Problems occur in evaluation learning program as described, ultimately drives the born of an alternative scoring methods of multiple choice test that can manage an answer (response) from test participants with more complex. Over time, and with the proper training in the use of the alternative scoring methods in multiple choice tests, can improve students' motivation and academic grades, because after all, the ultimate goal of the test is to provide adequate information for teachers and students that can be used in efforts to enhance the process of teaching and learning.

There are so many alternative scoring method of multiple choice test which has been introduced by experts, but this research will only focuses on the study of Formula Scoring (FS) method and the Right Number Elimination Testing (NRET) scoring method, where both scoring methods are equally applying the penalty system on the wrong answer. Penalty system is the traditional approach used in an attempt to reduce guesses on the answers to multiple choice test. The underlying logic of this approach is to prevent the test participants earn points that should not be accepted (Crocker & Algina, 1986: 399-400). A penalty system in scoring multiple choice test is expected to makes students to be more careful in answering the test and avoid guessing. Penalty here is a sentence reduction of the value of the score (Naga, 2013: 88). When students are already aware that guess would risk reducing their exam scores, then students will prepare better for the test and modify their learning strategies so that they can answer the test correctly without having to guess. Different with FS method, NRET is also equipped with a credit/ score on the partial ability of students in addition to a penalty system, so that NRET is able to detect the partial knowledge/ abilities and students' misconceptions.

The mathematical formula for the FS method is as follows (Frery, 1988: 33).

$$FS = R - \frac{W}{(C - 1)}$$

in which

FS = "corrected" or formula score

R = number of items answered right

W = number of items answered wrong

C = number of choices per item (same for all items)

Table below contains the test instruction and scoring guides for NRET. (Lau, et al, 2011: 100).

Table 2. NRET Test Instruction and Scoring Guides

Test Instruction	
<p>Must choose ONE option as the ANSWER by using "✓ CORRECT." ELIMINATE option(s) that you are SURE ARE NOT THE ANSWER by using "X WRONG." USE "? NOT SURE" if you are NOT SURE of an option. You have the flexibility to choose NONE (0), ONE (1), TWO (2), or THREE (3) "X WRONG," or "? NOT SURE."</p>	
Scoring Guides	
<p>ONE (1) point awarded if the option with "✓ CORRECT" is the correct answer. ONE (1) point awarded for each option eliminated correctly with "X WRONG". A PENALTY of 3 points deducted if the correct answer is eliminated with "X WRONG." Your score will range from -3 to 4.</p>	

Table 3. NRET Scoring Taxonomy

Score	NRET					
	FM	PM		AK	PK	
	-3	-2	-1	0	2	3
						4

Note: FM = full misconception, PM = Partial misconception, AK = absence of knowledge, PK = partial knowledge, and FK = full knowledge ((Lau, et al, 2011: 106).

Rules in multiple choice test using penalty system in its scoring methods is considered as something risky, because the students are asked to make decisions under uncertainty. In situations such tests, not only cognitive ability are shown students, but the level of students' risk

taking attitude can affect students' overall learning strategy that ultimately affect the study results. It is obvious that people differ in the way they resolve work-related or personal decisions that involve risk and uncertainty (Weber, Blais, Betz, 2002: 263). According to the tendencies to take a risk, people categorized as risk seeking, risk neutral and risk averse (Weber, Blais, Betz, 2002: 263-290). The relationship between the scoring method of multiple choice test and risk taking attitude toward chemistry learning outcomes is important for further investigation.

II. OBJECTIVES AND SIGNIFICANCE OF RESEARCH

A. Objectives

The general objective of this study was to determine: (1) the differences in chemistry learning outcomes among students who were given a multiple-choice test with Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, (2) the interaction between multiple choice test scoring method (Number Right Elimination testing and Formula Scoring) and students' risk taking attitude towards chemistry learning outcomes, (3) the differences in chemistry learning outcomes among students who categorized as risk seeking given a multiple choice test with Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, and (4) the differences in chemistry learning outcomes among students who categorized as risk averse given a multiple choice test with Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method.

B. Significance of Research

The results of this study are expected to be useful to: (1) provide insight to the teacher about the various scoring methods of multiple choice test that can be used in a formative evaluation in order to stimulate students' motivation to learn and improve student learning outcomes; (2) provide insight on risk taking attitude of students in relation to learning outcomes, especially in the using of various scoring method in multiple choice test; (3) serve as the basis for strategic decision-makers in education, especially the policies relating to the educational evaluation; (4) used as empirical foundation for future researchers, both in studying the evaluation of education as well as the characteristics of individual students.

III. RESEACH METHODOLOGY

A. Research Design

This research used experimental methods with design treatment by level (2x2). The treatment variable in this study is scoring methods on multiple choice tests, while attribute variable is a risk taking attitude that consists of two category, those are risk seeking and risk averse. The dependent variable in this study is chemistry learning outcomes.

Experiments carried out by treating such formative multiple choice test with NRET scoring method and FS scoring method is carried out systematically as much as four times in the past 12 meetings learning. Wherein each formative test done every sub topic completed by the teacher. After 12 meetings learning and four formative tests are completed, followed by an achievement test. To avoid bias in research it is necessary to control the internal validity and external validity of the experiment.

B. Population and Sample

The population in this study was 155 students of SMAN 13 Kota Bekasi in 11th (XI) grade of Natural Sciences (IPA) major. Research samples were taken by using simple random sampling technique. Successively conducted by: (1) choose two classes of the four classes XI IPA at SMAN 13 Kota Bekasi with the simple random sampling technique, using a lottery system (Gulo, 2005: 84); (2) The elected class given treatment formative multiple choice test with NRET scoring method and one other class treated with formative multiple choice test with FS scoring method ; (3) Conducting risk taking attitude tests simultaneously to all students in the two classes XI IPA elected to obtain information risk taking attitude of students, using risk taking attitude instrument; (4) The risk taking attitude scores of the two classes are sorted from largest to smallest; (5) students who have an average score on a risk taking attitude instrument of greater than 0.5 standard deviations above the average are categorized as risk seeking, while students who have an average score on a risk taking attitude instrument less than 0.5 standard deviations below the average are categorized as risk averse, the number of samples were selected in each cell in accordance with the study design is 10 students, so the total number of samples is 40 students.

C. Instruments and Data Collection Procedure

In this research, there are two kinds of data collected through the research instruments; those are the data of students' chemistry learning outcomes and data of students' risk taking attitude. Instrument of chemistry achievement test used to measure student learning outcomes after getting treatment. Chemistry achievement test developed by researchers, consists of 28 multiple choice items for chemistry topic of hydrolysis salt solution and buffer, internal consistency reliability of this instrument is 0.83. For the risk taking attitude instrument, researchers adapted from A Domain-specific Risk-attitude Scale (DOSPERS scale) developed by Weber, Blais and Betz (Weber, Blais, Betz, 2002: 263-290), the items in each domain were modified to measure students' risk taking attitude specifically on the situation of learning in schools especially learning chemistry, consists of 35 items with internal consistency reliability coefficient of 0.86.

D. Data Analysis Procedures

Normality of the data was tested using the Lilliefors test, while the homogeneity of variance was tested using Fisher and Bartlett tests. The hypothesis was tested using two way analysis of variance (ANOVA) followed with simple effect test using t-Dunnett test.

IV. RESULTS

The results of hypothesis testing using two way analysis of variance (ANOVA) can be seen in the following table:

Table 4. Result of Two Way Anova for Hypothesis testing

Source Variance	df	Sum of Squares	Mean Square	F	F _{crit}	
					0,05	0,01
Scoring Methods (A)	1	319,225	319,225	8,497	4,11	7,39
Risk Taking Attitudes (B)	1	0,025	-	-	-	-
Interaction (AxB)	1	1755,625	1755,625	46,730	4,11	7,39
Inter Group	3	2074,875	691,625	18,409	2,8	4,33
Within Group	36	1.352,500	37,56944	-	-	-
Reduction	39	3.427,375	-	-	-	-
Average	1	241025,625	-	-	-	-
Total	40	244453	-	-	-	-

As a consequence of the interaction, it is necessary to test the simple effect to examine the differences in chemistry learning outcomes among students who categorized as risk seeking given a multiple-choice test with NRET scoring method and FS scoring method, and the differences in chemistry learning outcomes among students who categorized as risk averse given a multiple-choice test with NRET scoring method and FS scoring method. T-Dunnett test results can be summarized in the following table:

Table 5. Summary of t-Dunnett Test

Comparison Group	t-Dunnett test		Conclusion
	t _{test}	t _{crit} (0,05)	
A ₁ B ₁ – A ₂ B ₁	6,784	1,729	Reject H ₀
A ₁ B ₂ – A ₂ B ₂	2,819	1,729	Reject H ₀

V. DISCUSSIONS

First, for the first research hypothesis, based on the results of two way ANOVA test, the value of $F_{test} = 8.497$, while the value of F_{crit} at significance $\alpha = 0.05$ is 4.11. So $F_{test} > F_{crit}$, thus H_0 is rejected. This means that there are significant differences in chemistry learning outcomes among students who were given a multiple-choice test with NRET scoring method and FS scoring method.

Based on the data collected, found that the average score of chemistry learning outcomes students who treated with multiple choice tests with NRET is higher than students who treated with multiple choice tests with FS. This finding is consistent with the different characteristics of the two scoring methods used in the study, even though both use penalty as correction for

guessing, but NRET also able to provide opportunities and credit/ score to the students' partial knowledge. Thus, the hypothesis put forward proved to be true, so the use of NRET scoring method in the multiple choice formative tests found to be more efficient in estimating the ability of students and to increase the effectiveness of feedback in the learning process, which in turn positively affects student learning outcomes.

Second, for the second research hypothesis, obtained $F_{\text{test}} = 46,73$. F_{crit} at significance level $\alpha = 0.05$ is 4,11. So $F_{\text{test}} > F_{\text{crit}}$, thus H_0 is rejected. This means that there are significant interaction between multiple choice test scoring method (NRET and FS) and students' risk taking attitude towards chemistry learning outcomes. This interaction can be interpreted as the effect of multiple choice scoring method on formative evaluation depend on students' risk taking attitude.

Each student has a different risk taking attitude, there are some students who tend to avoid risk (risk averse) and some are more likely to seek risk (risk seeking). Students who have an attitude as risk averse always felt uncomfortable in such uncertain conditions in the test situation by using penalty system. While students who have the attitude as risk seeking are able to offset risks in a more relaxed and inclined to challenge the uncertain situation with the potential. Therefore, students with different risk taking attitude respond to multiple choice scoring methods in a different way as well, so that the research data shows there are different learning outcomes for each treatment given. This is confirmed by the results of two way ANOVA test that showed a significant effect of interaction between multiple choice scoring methods and risk taking attitude towards chemistry learning outcomes.

Third, for the third research hypothesis, based on the result of simple effect analysis using t-Dunnet test, for a comparison the average score of chemistry learning outcomes in students who categorized as risk seeking given a multiple-choice test with NRET scoring method and FS scoring method, get the $t_{\text{test}} = 6,789$ while t_{crit} value at significance level $\alpha = 0,05$ is 1,729. So $t_{\text{test}} > t_{\text{crit}}$, thus H_0 is rejected. This means that there are differences in chemistry learning outcomes among students who categorized as risk seeking given a multiple-choice test with NRET scoring method and FS scoring method.

Students with a tendency to look for risk (risk seeking) remains a risk taker to guess the answers to multiple choice test although the test instructions has been said that the scoring rule implements a penalty for wrong answer, so guessing will not be tolerated. Students in this category looked at the FS scoring method more neutral and not seen as a threat. By contrast, when faced NRET scoring method, these students was facilitated and able to mobilize all the potential for them to maximize the results of tests. This is supported by research data showing that the average score in chemistry learning outcomes of the risk seeking students who treated with multiple choice tests with NRET was higher than who treated with multiple choice tests with FS.

Fourth, based on the result of simple effect analysis using t-Dunnet test, for a comparison the average score of chemistry learning outcomes in students who categorized as risk averse given a multiple-choice test with NRET scoring method and FS scoring method, get the $t_{\text{test}} = 2,819$ while t_{crit} at significance level $\alpha = 0.05$ is 1.729. So $t_{\text{test}} > t_{\text{crit}}$, thus H_0 is rejected. This means that there are differences in chemistry learning outcomes among students who categorized as risk averse given a multiple choice test with NRET scoring method and FS scoring method.

Students with category risk averse respond FS scoring method excessively, so these students trying to have maximal preparation before facing the test, because with the maximal preparation before tests made this group felt safe. Meanwhile, a system of penalties in NRET scoring method makes students with this category still feel threatened, so that these students less able to take advantage of the opportunity to maximize their score on NRET scoring method. On the other hand, the provision of credit/ score on the students' partial knowledge makes students with this category looked at the test situation using NRET safer than the test situation using FS, this view makes students less maximum in preparing for the exam. This is supported by research data showing that the average score in chemistry learning outcomes of the risk averse students who treated with multiple choice tests with NRET was lower than who treated with multiple choice tests with FS. However, differences in the average score of learning outcomes in the two treatments are not too large, this is because both scoring methods equally apply the penalty system on the wrong answer, it shows that risk averse students are more focused on a penalty system as a risk in both scoring methods rather than on the opportunities that can be gained from NRET scoring method.

VI. CONCLUTIONS

The result of this study revealed that: (1) there are significant differences in chemistry learning outcomes among students who were given a multiple-choice test with NRET scoring method and FS scoring method, it means that there is a significant effect of multiple choice scoring method toward chemistry learning outcomes. NRET scoring method found more effective and efficient to use in formative evaluation to enhance learning outcomes; (2) for risk seeking students, NRET scoring method gave a better result to maximize their learning outcomes than FS scoring method; (3) for risk averse students, FS provided higher learning outcomes than NRET; (4) effectiveness of multiple choice scoring methods on the formative evaluation were very dependent on the students' risk taking attitude, therefore both scoring methods can be used interchangeably in learning so that students with different risk taking attitude can still be facilitated by both.

REFERENCES

- [1] Middlecamp, C & E. Kean. *Panduan Belajar Kimia Dasar*. Jakarta: PT Gramedia, 1985.
- [2] Suryabrata, Sumadi. *Psikologi Pendidikan*. Jakarta: Raja Grafindo Persada, 1993.
- [3] Crocker, Linda and J. Algina. *Introduction to Clasicc al and Modern Test Theory*. Orlando: Holt, Rinehart and Winston, 1986.
- [4] Choppin, B. H. Correction for Guessing. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement: an international handbook*. Oxford: Pergamon Press, 1988.
- [5] Coombs, C. H., Miholland, J. E., & Womer, F. B. "The Assessment of Partial Knowledge." *Educational and Psychological Measurement*, Vol. 16, 1956.
- [6] Roediger, Henry L. & Elizabeth J. Marsh. "The Positive and Negative Consequences of Multiple-Choice Testing." *Journal of Experimental Psychology*, Vol. 31, No. 5, 2005.
- [7] Sudjana, Nana. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya, 2009.
- [8] Peraturan Pemerintah No. 74 tahun 2008 tentang Guru, Pasal 52, Ayat 2
- [9] Naga, Dali Santun. *Teori Sekor pada Pengukuran Mental*. Jakarta: PT. Nagarani Citrayasa, 2013.
- [10] Frary, Robert B. "Formula Scoring of Multiple Choice Test (Correction for Guessing)." *Instructional Topics in Educational Measurement*, Summer 1988.
- [11] Lau, Paul. N. K., et al. "Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests." *Educational Technology & Society*, 14 (4), 2011.
- [12] Weber, E.U., Blais, A.-R., & Betz, N. "A domain-Spesific Risk-Attitude Scale: Measuring Risk Perception and Risk Behavior." *Journal of Behavioral Decision Making*, Volume 15, 2002.
- [13] Gulo. W. *Metodologi Penelitian*. Jakarta: PT Gramedia, 2005.
- [14] Murray-Webster, Ruth dan David Hillson. *Understanding and Managing Risk Attitude*. Aldershot, UK: Gower, 2007.

DEVELOPMENT OF PERSONAL INTEGRITY SCALE: CONSTRUCT VALIDITY

Bambang Suryadi¹, Yunita Faella Nisa², Nenang Tati Sumiati³
^{1,2,3} Faculty of Psychology, Syarif Hidayatullah State Islamic University Jakarta
bambang.suryadi@uinjkt.ac.id

Abstract— The main purpose of this study is to develop personal integrity scale which measures integrity as personal traits. This study used mixed method combining both qualitative and quantitative methods. As for qualitative method, this study applied focus discussion (FGD) for collecting the data on the key words of integrity. Quantitative method was used in collecting the data of the study involving 442 respondents consisting 219 (49%) online respondents and 116 (51%) field study respondents. Integrity scale was used as an instrument of the study with 9 dimensions and each dimension has 30 items. Data analysis was done using Confirmatory Factor Analysis (CFA). The findings of the study indicate that integrity is simply defined as a alignment between individual's words and actions as characterized by honesty, sincerity, and consistency. The findings of the study also indicate that there are nine traits or attributes for personal integrity. They are honesty, keeping promise, loyalty, responsibility, persistence, kindness and caring, respect, fairness, and citizenship. Based on the definition of integrity and its nine personal traits or attributes, construct validity was applied resulting a personal integrity scale.

Keywords: *integrity, traits, attribute, honesty, loyalty, persistence, respect, and fairness.*

I. INTRODUCTION

In the 21st century, in any country, especially in developing countries like Indonesia, the issue of personal integrity becomes interesting issue. This is because personal integrity is not only a demand in the private sector and the company, but also in the government sector. In addition, personal integrity is also absolutely manifested in all aspects of life, ranging from social, political, social, economic, education, and culture. Without personal integrity, all sectors of life will collapse.

The results of literature review indicate that personal integrity is an important factor for the individual in carrying out his or her duties and responsibilities. A meta-analysis study conducted by Murphy and Lee (1994) found that there is a correlation between the integrity and the overall performance. This is in line with the study done by Ones, Viswesvaran and Schmidt (1993) which shows that integrity is a construct that includes responsibility, commitment to long-term work, consistency, things are prone to violence, moral reasoning, hostility, work ethic, dependency, depression and the level of energy. More specifically, this study measures the general construct of Conscientiousness (one of the big five personality dimensions).

Another research conducted by Tang and Liu (2011) showed that the higher the level of Authenticity of Supervisor's Personal Integrity and Character (ASPIRE) the higher the Love of Money (LOM), and self-esteem, but the lower Personal Unethical Behaviour (PUB). PUB significantly correlated machiavellianism high level, low self-esteem, and intrinsic low religiosity. The results of the study also indicate that the effect of LOM on PUB is not significant, but the influence of PUB on ASPIRE is significant. People with high LOM and low ASPIRE tend to have high PUB.

Based on the above-mentioned two studies, it can be concluded that personal integrity problems associated with ethical behavior or unethical behavior. Ethical or unethical behavior is a picture of the level of personal integrity. Individuals with a high level of integrity tend to demonstrate ethical behavior such as honesty, trustworthiness, and discipline. On the other hand, individuals with low integrity level tend to show unethical behavior, such as dishonesty, greediness, corruption, and noncompliance with regulations or unlawfulness.

It is important to note that the problems of unethical behavior is not only happening in Indonesia, but also in other countries. A Transparency International survey in 2012 showed that Indonesia was ranked at 118 position out of 176 countries. Together with Ecuador, Egypt and Madagascar, Indonesia is ranked 118 with a CPI score of 3.2 (scale of 0-10). While the first position in the world's cleanest countries filled by Denmark, Finland and New Zealand. These countries lead with a score of Corruption Perception Index (CPI) amounted to 9 on a scale of 0-10. With these results, the country can be considered as a country that is very clean, almost close to perfect.

In addition, the cases of unethical behavior also occurs in many other aspects of life. In economic field, the case is very prominent when the massive fraud committed by Enron, Worldcom and several other large companies were revealed in mid 2000 to 2002. The company committed fraud on how to manipulate the financial statements and in cooperation with public accounting firms to continue to provide a reasonable opinion on the financial statements which have been modified (Lasantu, 2012). In the area of governance, as many as 311 of the 530 heads of regions in Indonesia entangled case law, 86 percent of whom corruption cases (Setia, 2013). Plagiarism (Abimanyu, 2014; Yulianti, 2013) or harassment (Poet, 2014) among lecturers also be evidence of unethical behavior in the field of education.

Based on the above facts, it can be concluded that corrupt behavior as unethical behavior has been a latent disease in Indonesia and occurs in almost all public and private sectors. This corrupt behavior needs to be prevented early through preventive measures, if the Indonesian people want to be a dignified nation to uphold moral values and ethics. One of the strategic measures of a preventive nature to prevent corrupt behavior is by measuring the level of integrity of the person prior to being employed as public officials.

The problem is that until now, based on the results of the literature review, in Indonesia there is no instrument or standard measurement tool to measure the level of personal integrity. One of the tools that are often used in knowing the level of individual's integrity (honesty) is a lie detector machine. However, the use of this measure is debatable even doubt the accuracy and validity. In addition, practitioners in the field of assessment and selection is usually using Behavior Event Interview (BEI) to explore the tendency of one's integrity (Rahman, 2006). This method has disadvantages in term of limited number of sample size as so the findings are less accurate.

Recognizing the importance of personal integrity in the individual and social life and the lack of measurement tools or instruments standard to determine the level of one's personal integrity, the need for developing personal integrity scale becomes an urgent need. This measure would have predictive ability against someone in a variety of situations to carry out its duties and responsibilities, whether someone has the integrity level of low, medium, or high.

II. RESEARCH PROBLEM

Based on the background of the study as described above, the problem of this study can be formulated as follows: "What kind of instrument that can measure the type of traits that one should have to be an individual with a high level of personal integrity?"

III. LITERATURE REVIEW

A. Definition of Integrity

Many definitions of personal integrity have been given by experts in the field of psychology. However, the existing definitions is not conclusive, there for the writer needs to redefine the meaning of personal integrity. The word integrity comes from the English language. In English Thesaurus (UK), the term integrity has some synonym such as honesty, truthfulness, honor, veracity, trustworthiness, and sincerity. The opposite of integrity is dishonesty.

According to Robert (1999) "the word integrity means "wholeness," wholeness of virtue, wholeness as a person, wholeness in the sense of being an integral part of something larger than the person—the community, the corporation, society, humanity, the cosmos".

Another interesting definition of integrity is given by Becker (2009) who said that personal integrity is consistency between one's personal principles and moral respect to people. This requires both a commitment to establish ourselves as moral subjects (like what we should be)

and to let the moral autonomy of others. Furthermore Becker (2009) asserts that personal integrity is a moral obligation to respect others as human and moral standards to everyone to judge whether a person is immoral or not.

From the above discussion, it can be understood that the term integrity has multiple keywords, such as honesty, real situation, sincerity, accuracy, reliability, perfection, harmony, safety, and consistency. Given these keywords, in the current study the personal integrity can be defined as individual's consistency between his or her thoughts, words and actions characterized by honesty and sincerity.

B. Integrity Dimensions

The results of a literature review on the dimensions of personal integrity over the last decade (2005-2015) indicate that there have been no conclusive conclusions related to the number of personal integrity dimension. In other words, there are perspectives in the dimensions of integrity as given by experts. This dimensional variation indicates that the terms of personal integrity has its own complexity. The following sections will explain some of these variations.

ICAEW (2007) identified five aspects of integrity, which includes moral values, motives, commitment, quality and achievement. Unlike the ICAEW (2007), Barnard, Schurink, & Beer (2008), identified 10 aspects of personal integrity which include, self-motivation and encouragement, moral courage and decisiveness, honesty, consistency, commitment, diligence, self-discipline, responsibility, trust, and fairness.

A recent study by Baxter, Dempsey, Megone, & Lee (2012) shows that personal integrity consists of four aspects as follow:

1. Wholeness of character: consistency between individual's action and words;
2. Ethical value: individual's action on task commitment;
3. Identity: Individual's definition on ethical commitments; and
4. Standing for something: Individual's values.

Taking into account the variations in the dimensions of personal integrity as described above, in this study the dimensions of personal integrity as the focus of this study are as follows:

1. honesty,
2. keeping promises,
3. loyalty,
4. responsibility,
5. persistence,
6. kindness and caring,
7. respect,
8. fairness, and
9. citizenship.

C. Factors Affecting Integrity

According to Paaenen (in Murphy, 1993) personal integrity in relation to dishonesty, not only influenced by individual factors, but also influenced by the factor of the situation. He identified a number of individual variables considered as factors affecting dishonesty. These factors are:

1. undependability, the individual irresponsible, impulsive, careless, and so are considered to be associated with dishonesty and fraud,
2. problems in socialization, individuals with values that have not evolved, or has a history of delinquency,
3. attitudes regarding deviance and theft, individual who have a positive attitude towards thieves and delinquent behavior (naughty).
4. problems with authority relationships, individuals who have difficulty for interaction or relationship with the authorities.
5. excitement seeking, individuals who engage in thrill-seeking behavior or other courageous.
6. work motivation, individual with low level of work motivation.
7. social influences; individuals are easily influenced by a friend of the same age or others
8. unstable upbringing, individuals with unstable family life.

9. drug use, individuals with a history of drug abuse.
10. unmet need, individuals with low self-esteem, low job satisfaction, and so on associated with dishonesty and fraud.

IV. RESEARCH METHOD

This study used a mixed research method, i.e. combining a quantitative and a qualitative method. The former is used in order to get the respondents' opinions on the dimensions and indicators of personal integrity. The latter is used in order to deepen and widen the research problems and it was done through Focused Group Discussion (FGD) involving certain figures such as academicians, professionals, politicians, and businessman.

The respondents of this study were 442 people, consisted of 216 (49%) online survey respondents and 226 (51%) field study respondents. The first category of respondents consisted of people ages between 19 to 50 years, such as teachers, lecturers, professionals, house wives, and students. Their educational background also varied, ranging from senior secondary school to tertiary level with Ph.D holder (S3).

As for the field study respondents, it consisted of 226 people with 60 (27%) male and 166 (73%) female, aged between 21 to 24 years. They were students of Faculty of Psychology UIN Jakarta.

In collecting the data, this study used Personal Integrity Scale developed by the researchers with reference to the theories and concepts of personal integrity. It has nine parts in line with the personal integrity aspects as explained in the conceptual framework of this study. Each dimension has 30 items, making a total of 270 items for the whole instrument. The items were written in both favorable and unfavorable statements.

The instrument was developed in a Likert Scale Model with four different responses for each statement (item). For items measuring respondents' cognitive response, the response options were given from Strongly Disagree, Disagree, Agree, and Strongly Agree. For items measuring respondents' behavior, the response options were given from Very Unsuitable, Unsuitable, Suitable, and Very Suitable. Finally, for items measuring respondents' frequency response, the response options were given from Never, Some times, Often, Very Often.

V. RESEARCH FINDINGS AND DISCUSSION

In general, the findings of the study indicate that personal integrity has a broad meaning and a varied definition. Given the key words, definitions, items, and meanings gained from FGD, online survey, and field study, the researchers come up with the conclusion that personal integrity is simply defined as individual's consistency between his or her thoughts, words and actions characterized by honesty and sincerity.

More specifically, the findings of this study indicate that personal integrity has nine dimensions as follows:

- a. honesty,
- b. keeping promises,
- c. loyalty,
- d. responsibility,
- e. persistence,
- f. kindness and caring,
- g. respect,
- h. fairness, and
- i. citizenship.

Given the above mentioned definition and nine dimensions of personal integrity, this study has produced a Personal Integrity Scale with 152 items constructed in both favorable and unfavorable statements. Below is the sample of items for Personal Integrity Scale.

Table 1. Sample of Items for Personal Integrity Scale

Aspect	Total item	Item
Honesty	14	I left work early without permission even though it is cheating
Keeping Promises	21	I am careful to say promises
Loyalty	28	I recommend the company as a good place to build a career
Responsibility	12	I think of every action so that I can give an account
Persistence	17	I know what my purpose in life
Kindness And Caring	16	In my opinion, showing concern for others is the right thing
Respect	17	I was able to understand my mind.
Fairness	5	I never expect a reward for what I did.
Citizenship	22	I was recognized as part of a community.

As stated in the conclusions of this study, the term personal integrity has a very broad meaning and significance. Therefore, in this study, to obtain conclusive definitions of personal integrity, the researchers carried out in-depth literature review, focus group discussions (FGD), and an online survey. Based on the data gathered in this study, the researchers defined as individual's consistency between his or her thoughts, words and actions characterized by honesty and sincerity. This definition needs to be studied further so that it can be widely accepted by many people and professions. In this context, the researchers want to discuss how this research could come to such a conclusion.

It is important to note that consistency between the individual's words and actions has become a vital element for defining personal integrity. From the religious perspective, inconsistency between individual's words and action is called *nifaq* or hypocrisy. The word *munafiq* is derived from *nifaq* which has three characteristics, namely, being a liar when he or she said something, denying the promise he or she has made, and not keeping the trust given to him or her. The three traits of *munafiq* are very dangerous for individuals and for others.

In short it can be concluded there are three characteristics or traits that became the main feature for individuals to have personal integrity. They are honesty, sincerity and consistency. One question that arises is whether these three characteristics or traits show the hierarchy or not? If showing a hierarchy, there may be those who say that honesty or sincerity should be in the first position, so the order becomes sincerity, honesty, and consistency. In this respect, sincerity becomes the foundation of all actions and behavior of individuals. Without sincerity, all the actions and deeds of individuals will be meaningless.

After sincerity, the second attribute associated to personal integrity is honesty. Honesty is simply defined as being genuine and objective in all aspects of individual behavior. Honest people tend to show good conduct and behavior, while dishonest people tend to cheat and misbehave. In the modern life, honesty has become very rare attribute found in individual's characteristics. Therefore honesty is a major asset in achieving one's success. Honesty is the best policy.

The third trait is consistency. The word consistency in religious perspective is called *istiqomah* which means steady and unchanging behavior under any circumstances. In this context, the individual's sincerity and honesty should be done consistently.

Another aspect that needs to be addressed in this study is the nine dimensions of personal integrity, which include honesty, keeping promise, loyalty, responsibility, persistence, kindness and caring, respect, fairness and citizenship. Given these characteristics, one will most likely ask a question: How this research could come to the conclusion that personal integrity consist of nine dimensions? To what extent the relationship between one dimension and another dimension? Is every dimension really unidimensional or multidimensional?

The answers to these questions may be given through analytical techniques used in this study, that is the exploratory factor analysis (EFA). With EFA researchers could detect the structure within the relationships between variables, so that we can classify the variables. In

addition, using EFA, we are able to reduce the big number of variables into a relatively smaller variables. However, this kind of analysis needs to be performed repeatedly with a large sample to make the result valid and accurate.

VI. RECOMMENDATIONS

Based on the research findings as presented in the previous chapter, in this section, the researchers would like to give both theoretical and practical recommendations.

A. Theoretical Recommendations

Since this study is a preliminary study which aims to develop a measuring tool of personal integrity and considering the limitations of the present study, the researchers give the following theoretical recommendations.

1. Further study needs to be done in a comprehensive way to formulate the concept of integrity in the context of Indonesia. The study should be conducted using two approaches, namely the psychological and religious approach.
2. The need for further study to explore the local wisdom associated with personal integrity. Local knowledge can be formulated in terms of the properties that can describe the high level of integrity for individuals with attention to cultural and ethnical diversity in Indonesia.

B. Practical recommendations

For researchers interested in undertaking further research, the following practical recommendations might be taken into account.

1. The present study is limited to the study of integrity as an individuals trait, therefore further research is needed to examine the integrity of a conditional state that makes someone show personal integrity.
2. Given the sample is limited to students, therefore further research is needed with a broader sample involving several elements, for example, samples of teachers, lecturers, government or private employees, professionals, businessmen, and politicians.
3. Need to do further research using the scale of personal integrity that have been developed to measure the level of employee integrity of an institution. For the initial stage, a further study can be done internally at the State Islamic University (UIN) Syarif Hidayatullah Jakarta.

REFERENCES

- [1] Anggito Abimanyu resigned from UGM after being accused of plagiarism (17 Feb 2014). Retrieved from antaranews.com tanggal 29 March 2014.
- [2] Baxter, J., Dempsey, J., Megone, C., & Lee, J. (2012). Real Integrity: practical solution for organizations seeking to promote and encourage integrity. London: ICAEW.
- [3] Becker, G. K. (2009). Integrity As Moral Ideal And Business Benchmark. *Journal of International Business Ethics*. 2 (2), 70-84
- [4] Lasantu, A.I. (2012). *Conditions of Corruption in Government and Private Sector and Handling in the US*
- [5] Murphy, K.R. (1994). *Honesty in the Workplace*. Brooks/Core Publishing Company : United States America
- [6] Ones, D., Viswesvaran, C., & Schmidt, F.L. (1993). Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *The American Psychological Association*, 78, 679-703.
- [7] Ones, D.S., Viswesvaran, C & Schmidt F.L. (1993). Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *Journal of Applied Psychology*. 78(4), 679-703.
- [8] Rahman, E. (2006). Identifying Integrity, It is possible? (Meraba integritas, bisakah?). Kompas. Excerpt, Jakarta, 2006.
- [9] Setia, L. (2013). The election and the corruption party of District Leader. Retrieved from daerah.sindonews.com on 29 March 2014
- [10] Tang, T.L., & Liu, H. (2012). Love of Money and Unethical Behavior Intention: Does an Authentic Supervisor's Personal Integrity and Character (ASPIRE) Make a Difference?. *Journal Business Ethics*. 10(7), 295-312. DOI 10.1007/s10551-011-1040-5
- [11] The Institute of Chartered Accountants in England and Wales (ICAEW). (2007). Reporting With Integrity Information For Better Markets Initiative. ICAEW: England.

Argument-based Validity of Situational Judgment Test for Assessing Teaching Aptitude

Budi Manfaat

Departement of Mathematics Education, IAIN Syekh Nurjati Cirebon

budi_manfaat@yahoo.com

Abstract— Assessment of teaching aptitude required for selection-decision making in admission to teacher education program. In conventional admission procedures, teaching aptitude assessed explicitly through self-report or interviews. The problem with self-report in high-stakes situations are respondents can potentially fake their responses and portray themselves as best as possible. The problem with interviews in large-scale situation are expensive and requires a long time. To overcome these problems, teaching aptitude assessment can be done implicitly through the *Situational Judgment Test* (SJT). Through SJT, teaching aptitude can be indirectly by asking individual to assess the effectiveness of responses to situations that is designed to obtain targeted traits. Teaching aptitude domains involves commitment, communication, and empathy. This study aims to develop a valid SJT blueprint to assess teaching aptitude. Validation is done by using Kane's model, that is argument-based validation. In Kane's model, the validation process consists of two phases: an *interpretive argument* and a *validity argument*. In the interpretive argument, the proposed uses and interpretations of SJT score are specified and assumptions being made are explicitly identified. In the validity argument, logical and empirical evidence is gathered to evaluate the plausibility of these inferences and assumptions and of the reasoning that connect these elements. The results of this study will help test-developer constructing a SJT for admission to teacher education program.

Keywords: teacher, teaching aptitude, SJT, validity argument

I. INTRODUCTION

Mission of the teacher education program is development process from teaching aptitude to teaching competencies. Cronbach and Snow [1] proposed Aptitude-Treatment Interaction (ATI) theory, suggests that optimal learning results when the instruction is exactly matched to the aptitudes of the learner. It is consistent with Gardner's Theory of Multiple Intelligences [2] that suggest a multidimensional view of ability. According to Gardner, the implication of the theory is that learning should focus on the particular intelligences of each person.

Aptitude defined as "any characteristic of a person that forecasts his probability of success under a given treatment" [1]. Definition of aptitude expanded from the conventional cognitive-based strategies and abilities, to include conative and affective characteristics [3]. Conative being motivational and volitional aspects of learning and affective being temperamental and emotional aspects. Aptitudes are natural abilities which must be developed to become competencies through the systematic learning and training [4]. The goal of teacher education program is to develop teaching competencies. According [5], there are four competencies for good teaching, they are: content knowledge, pedagogical knowledge, pedagogical skills, and attitudes. For effective teaching practices, an essential domain of pedagogical skills is communication skill, whereas an essential domain of attitudes are commitment and empathy. Commitment is conative characteristic, whereas empathy is affective characteristic. Therefore, teaching aptitude then defined as any characteristic of a person that forecasts his probability of success under teacher education program.

In conventional admission procedures, teaching aptitude assessed explicitly through self-report or interviews. The problem with self-report in high-stakes situations are respondents can potentially fake their responses and portray themselves as best as possible [6]. The problem with interviews in large-scale situation are expensive and requires a long time. To reduce these problems, teaching aptitude assessment can be done implicitly through the *Situational Judgment Test* (SJT) [7]. Through SJT, teaching aptitude can be indirectly by asking individual to assess

the effectiveness of responses to situations that is designed to obtain targeted traits. For simply, in this article, SJT for assessing teaching aptitude shorted by SJT-TA.

As a high-stakes testing, validity of the use and interpretation of SJT-TA scores must be shown. Test validation is the process of making a case for the proposed interpretation and uses of test scores. This case takes the form of an argument that states a series of propositions supporting the proposed interpretation and uses of test scores, and summarizes the evidence supporting these propositions [8]. The purpose of this study was to develop a valid SJT-TA blueprint.

II. RESEARCH METHOD

Proposed SJT-TA includes three non-cognitive domains: commitment, communication, and empathy. The indicators of each domain are presented in Appendix 1. For each domain includes a series of dilemmatics situations. For each dilemmatic situation, testees asked to rank-order five responses, from the most appropriate to least appropriate action. Example of SJT-TA item presented in Appendix 2.

Validity of the proposed SJT-TA will be shown by using Kane's model of argument-based validation. In Kane's model, the validation process consists of two phases: an *interpretive argument* and a *validity argument*. In the interpretive argument, the proposed uses and interpretations of SJT-TA score are specified and assumptions being made are explicitly identified. In the validity argument, logical and empirical evidence is gathered to evaluate the plausibility of these inferences and assumptions and of the reasoning that connect these elements.

Development of such arguments requires the use of a clear structure on which the argument may be based. According to Toulmin's argument structure [9], argument structure is built on several components, which include the grounds, claim, warrant, backing, and rebuttal. The claim (or proposition) of an argument is the conclusion one draws about an individual based on test performance whereas the grounds serve as the data or observations upon which the claim is based upon. The inference linking the grounds to the claim is not given and therefore justification is needed in the form of a warrant (or assumption). The warrant in Toulmin's model is considered to be a rule, principle, or inference-license that is meant to provide justification for the inference connecting the grounds to the claim. Warrants in turn need backing (or evidence) which comes in the form of theories, research, data, and experience.

To establish a connection between the claims and grounds, three types of inferences must be used in a chain to connect observations and conclusions [10]. These three inferences were identified as scoring, generalization and extrapolation. Scoring is an inference from an observation of performance to a score, and is based on the assumptions about the appropriateness and consistency of the scoring procedures and the conditions under which the performance is obtained. Generalization is an inference from an observed score on a particular measure to a universe score, or the score that might be obtained from performances on multiple tasks similar to those included in the assessment. This inference is based on the assumptions of measurement theory. Extrapolation is an inference from the universe score to a target score, which is essentially an interpretation of what a test taker knows or can do, based on the universe score. This link relies on the claims in an interpretative argument and the evidence supporting these claims.

The proposed used of SJT-TA is to assess teaching aptitudes in applicant to teacher education program, and the interpretation is that higher scores would indicate better teaching aptitudes therefore match to learning on teacher education program. Propositions that underline the proposed test score interpretation and uses, the assumptions, and the evidence relevant to each proposition are presented in Appendix 3. Plausibility of the interpretive argument would be based on the relationship between data collection procedures and the proposed interpretation. For gathering empirical evidence, the participants were 130 students of Mathematics teacher education program in IAIN Syekh Nurjati Cirebon selected.

III. RESEARCH FINDING AND DISCUSSIONS

Assumptions (1) and (2) imply that performance on the SJT-TA is relevant to readiness for learning in teacher education program, at least in the sense that applicants with low scores on the test are likely to have difficulty in developing teaching competencies, because they lack of teaching aptitudes. Furthermore, assumption (3) can

draw the stronger conclusion that applicants who pass the test (score at or above the cutoff) are "ready" for the teacher education program (acceptable) and that applicant who fail the test are not ready for the teacher education program (rejectable). However, applicants who fail the test can also be accepted with a certain consequence (eg by giving preferential treatment in the program). This argument is in line with the ATI Theory.

The importance of non-cognitive attributes (commitment, communication, and empathy) for effective teaching has been shown and suggested by a number of researchers. While, more teacher education programs focus on developing content and pedagogical knowledge. Thus, admissions to the teacher education program based on SJT-TA scores is rational decision.

The average difficulty and discriminant index for SJT-TA domain are presented in Tabel 1. The items with difficulty index between 0.60 and 0.90, and discrimination index above 0.20, show highly discriminating item with moderate difficulty, typical of good items [11]. Split-half reliability coefficient is 0.74, show high reliability. While, coefficient of correlation between SJT-TA scores and rating score of student performance in micro-teaching by expert-rater is 0.64, show high positive relationship between teaching aptitude (as predictor) and teaching performance (as criteria). This empirical evidences supporting the proposed scoring, generalization, and extrapolation inferences.

Table 1. Average item difficulty and item discrimination index of SJT-TA based on the responses of 130 participants

Item Analysis	Domain		
	Commitment	Communication	Empathy
Difficulty Index	0.70	0.71	0.66
Discrimination Index	0.41	0.29	0.33

IV. CONCLUSIONS AND SUGESTIONS

Logical (theory-based) and empirical evidences collected in this study supporting the uses and interpretations of SJT-TA that proposed. Therefore, the results of this study can help administrators to constructing a SJT-TA items for applicant screening or selection into teacher education program. However, several limitations of the empirical evidences should be considered. Ideally, empirical evidences base on big-data from vary location and condition. Furthermore, ideally, score of SJT-TA gathered from applicants (not students) at teacher education program. Thus, further investigation to strengthen support empirical evidence needs to be done.

REFERENCES

- [1] L.J., Cronbach & R.E., Snow, Aptitudes and instructional methods, New York: Irvington, 1977
- [2] H., Gardner, Frames of mind: The theory of multiple intelligences, New York: Basic Books, 1983
- [3] Snow, R.E. 1995
- [4] F., Gagne, From giftedness to talent: a developmental model and its impact on the language of the field, Roeper Review, 18(2), 1995, pp. 103-111
- [5] C.E., Casey & R.A., Childs, Teacher education program admission criteria and what beginning teachers need to know to be successful teachers, Canadian Journal of Educational Administration and Policy, 67, 2007, pp. 1-18.
- [6] R.M., Klassen, E., Rowett, T.L., Durksen, & F., Patterson, Applicant reactions to a situational judgment test used for selection into initial teacher training, International Journal of Educational Psychology, 3 (2) 2014, pp. 104-124.
- [7] D.L., Whetzel & M.A., McDaniel, Situational judgment tests: An overview of current research. Human Resource Management Review, 19, 2009, pp. 188-202
- [8] M.T., Kane, Validation. In R. L. Brennan (Ed.) Educational Measurement (4th ed.). Westport, CT: American Council on Education and Praeger, 2006
- [9] S.E., Toulmin, The uses of argument: Updated edition. Cambridge, UK: Cambridge University Press, 2003
- [10] M.T., Kane, T.J., Crooks, & A.S., Cohen, Validating measures of performance. Educational Measurement: Issues and Practice, 18, 1999, pp. 5-17
- [11] T.M., Haladyna, Writing Test Items to Evaluate Higher Order Thinking. USA: Allyn and Bacon, 1997

Appendix 1: Domains and elements of SJT-TA

Commitment to Teaching Profession

- Enthusiastic about subject matter
- Enthusiastic about helping other people to grow and develop;
- Desires to be a good teacher
- Consistent line of activity
- Resilience (ability to navigate the rocky road between vision and reality)

Effective Communication

- Listener (attentive and not dominant)
- Engagement
- Humour

Empathy

- Sensitivity
 - Patience
 - Respect
 - Tolerance
 - Acceptance
 - Understanding
 - Flexibility
 - Openness
-

Appendix 2: Example of SJT-TA item

Situation:

You have a twin younger brother, say their name is Adin and Ale. A few more days they will take a national exam. You as a brother, prompted by your parents to help them. In terms of motivation to learn, Adin much different with Ale. Adin is very easy to learn, while Ale opposite. Ale is easier to play than learning.

Rank in order the following actions in response to this situation (1= Most appropriate; 5= Least appropriate)

- A. Ale does not need to be prosecuted to learn.
- B. Need to flaunt the superiority of Adin on Ale, so Ale motivated to learn.
- C. Try a variety of ways so that Ale want to learn like Adin.
- D. Accompanied Ale to play, then slowly invites him to learn.
- E. Forcing Ale with a hint of a threat, and if it doesn't work, I will leave him.

Scoring Key:

Ideal Rank	Applicant Rank 1	Applicant Rank 2	Applicant Rank 3	Applicant Rank 4	Applicant Rank 5
D	4	3	2	1	0
C	3	4	3	2	1
B	2	3	4	3	2
E	1	2	3	4	3
A	0	1	2	3	4

Appendix 3: Propositions that underline the proposed test score interpretation and uses, the assumptions, and the evidence relevant

Inferences	Proposition (Claim)	Assumptions (Warrant)	Evidence (Backing)
Scoring	Task and scoring criteria of SJT-TA are appropriate for obtaining evidence of readiness applicant to learning in teacher education program. [Proposition (1)]	Teaching aptitude needed in order to be successful in teacher education program. [Assumption (1)]	Aptitude-Treatment Interaction (ATI) theory; Multiple Intelligence Theory; Implicit Trait Policy (ITP) Theory. Systematic development of rubrics for scoring; Item Analysis
Generalization	Observed SJT-TA scores are consistent over future with parallel task versions so that expected scores can be estimated. [Proposition (2)]	Teaching aptitude measures required in teacher education program, is reasonably reliable, and is not influenced substantially by any sources of systematic error. [Assumption (2)]	Behavioral Consistency Theory; Reliability statistic (split-half reliability)
Extrapolation	Performance on the SJT-TA is related to criteria of success study in teacher education program. [Proposition (3)]	The cutoff score is appropriate in the sense that applicants with scores at or above the cutoff score have sufficient teaching aptitude to succeed in teacher education program, and applicants who score below the cutoff lack some or all of the needed for the calculus course. [Assumption (3)]	Relationship between SJT-TA score and performance on micro-teaching practice.

Horizontal Equating in Accounting Vocational Theory Test Based on Mean/Mean Method of Item Response Theory

Dian Normalitasari Purnama¹, Sigit Santoso²

¹ Student of Educational Research and Evaluation, Yogyakarta State University

² Student of Educational Research and Evaluation, Yogyakarta State University

diannsp@gmail.com

Abstract--This research is a descriptive quantitative research which aims to determine the equation of the horizontal equating of accounting test between test A and test B according to Mean/Mean method in Item Response Theory approach. The research is carried on two tryout test packages of accounting vocational theory made by MGMP Sleman in academic year 2015/2016. The source of data obtained from the answer sheets of students who received test A and test B, respectively of 325 sheets. The samples are taken by using stratified random sampling. Item characteristic is seen by Item Response Theory approach using *Bilog* 3.0 program, while equating process is performed by Mean/Mean method assistance of Microsoft Excel program. The analysis shows that the characteristic of test A has 27 item categorized as good characteristic and suited for the 2 PL model, meanwhile the characteristic of test B has 24 item categorized as good characteristic and suited for the 2 PL model. The result of the equating test using Mean/Mean method generates equation $b_2^* = 0,997bx - 0,250$.

Keywords: *Accountancy, Characteristic, Equating, IRT, Test*

I. INTRODUCTION

Assessment and evaluation of students' learning outcomes is one of the important things to do by educators and educational institutions. Assessment and evaluation of learning outcomes is an effort which undertaken to monitor the achievement of student competencies after learning process. According to Law No. 20 of 2003 article 57 clause 1 about national education standard, evaluation is held in order to control the national education quality as a form of accountability of education providers to the stakeholders, including the teachers-students, institutions, and educational programs. One of the government's evaluations toward the implementation of education is to conduct the national examination.

National examination is conducted by government once a year simultaneously. Dealing with Educational Ministerial Decree No. 20 of 2007 about assessment standard is stated that national exam is measurement activity against students' competencies in several subject base in the knowledge and technology in order to assess the achievement of national exam. The results of national examination are then used by the government to make policies, particularly in the field of education. Based on Government Regulation No. 19 of 2005 article 68 about national education standard is stated that the results of national examination are used as one of considerations for mapping the quality of programs and education units. The mapping aims to look at the quality of education in each region of Indonesia.

The Province Educational Authority has always held the trial test in dealing with national examination, commonly referred as tryout, before the national examination is conducted. The aim of tryout is to enhance the quality of students' graduate. This is done to see the attainment of student competence in the implementation of the national examination, as well as the step of data validation and preparation for students dealing with national examination. The importance of tryout implementation dealing with national examination requires The Educational Authorities as the organizers to prepare tryout very well.

The planning of tryout is carried out in steps. One of the most important steps is item test preparation. The implementation of tryout requires test instrument which has good quality. The tryout's result would show the actual of student achievement competency if the test instrument used in tryout was in good quality. In addition, the tryout's result could be useful to analyze the competence of which students have not been achieve yet.

The interview which was conducted by researchers to accountancy teacher who is also the chairman of MGMP in accountancy in *Slleman*, show that the tryout test of accountancy in vocational high school arranged by MGMP. The test consists of two packages, which were arranged by using one blueprint and one range of material. The test is created into two packages, in order to avoid cheating during the tryout.

Relating to the two packages of test which is used in tryout of vocational accountancy test in *Slleman*, the main concern is how those two test packages could be used as a set of items with good item characteristic. Besides, the parallelism of those two test packages is unknown. That means the item level difficulty or item differential from both packages is not in the similar level. This could lead the students with the lower ability gain the high scores while the students with the high ability obtain the low score due to the item difficulty of the test is not equal.

Those situations could lead the measurement error which means that the implementation of tryout test could not show the achievement of students' competence that will be used as consideration in an effort to face the national examination. The measurement error leads the implementation of the tryout misses the target. Measurement error due to un-parallelism test results the improper constant conversion in both test packages. Hence, the parallelism of two test packages must be observed despite the blueprint and the material come from the similar source.

If the both test packages have been proven parallel, the next step is dealing with equating process. Sukirno (2007) explained that the equating process is able to compare students' score through different test packages, so the students' score will fair either they obtain the harder test or they obtain the easiest one. The existence of the equal measurement in students' score will ease the educational quality control. The test result will show students' competence in order to prepare the national examination, and also it can be used as a consideration for the decision to improve the quality of graduates.

The both test packages will have the equal measurement, if both of them have been equalized, which means that both of test packages can be compared one another. The equal measurement will also help to determine the competences which have been mastered by the students either obtaining in test packages A or obtaining in test packages B. Those test's result can be used as a consideration for the decision, and an improvement action dealing with national examination. Besides, if the item characteristic of the test has been known, the test can be used as the input of item bank development.

The explanation above shows the overview about the importance of equating process between two test packages. Analyzing and equating the test packages will give the valid information from the test instrument and will show the competency achievement in facing the national examination. Accordingly, the aim of this research is to determine the equation from the equation process in order to make the equal measurement.

II. METHODOLOGY

This research is a descriptive quantitative for equating two test packages of tryout test in accounting vocational theory made by MGMP in *Slleman*. The population of this research is students' responses which are obtained from the students' answer sheets of tryout test in accounting vocational theory, both of the packages A and packages B. The determination of the packages which would be undertaken by students was randomly chosen based on the students' test number; odd number obtained test packages A, and even number obtained test packages B. Kolen & Brenaan (2004; 15) stated that there are two ways to choose the sample in equivalent groups. The first is by giving the single test to identify the ability, while the second one is by dividing the test kit structurally, e.g. the first students obtained the test package X, second student obtained the test package Y, the third student obtained the test packages X, etc. Therefore, the students' ability in tryout test in accounting vocational theory is assumed equivalent.

The sample of this research use stratified random sampling. This technique results the level of the vocational high school based on the national examination scores which then categorized as high level, moderate level, and low level.

The data are obtained from students' answer sheets in tryout test. The equivalency of both test packages was done before equating process. This equivalency test aims to see whether the both test packages parallel or not. If the result of equivalency test showed that both test packages are not parallel, so the equivalency test should be done, *vice versa*. The method used

in this equivalency test is Mean/Mean method. The equating process is carried by using equivalent group design.

The equating method, Mean/Mean method, involves two parameters, i.e. discrimination parameter and difficulty parameter. Kolen & Brennan (2004: 162) stated that the relationship between item parameter estimation from both equating test based on the equation as follow:

$$b_2 = \alpha b_1 + \beta \dots\dots\dots (1)$$

$$a_2 = \frac{a_1}{\alpha} \dots\dots\dots (2)$$

$$\text{Hence, obtained } \bar{b}_2 = \alpha \bar{b}_1 + \beta$$

$$\text{So, } \alpha = \frac{\bar{a}_1}{\bar{a}_2} \dots\dots\dots (3)$$

$$\beta = \bar{b}_2 - \alpha \bar{b}_1 \dots\dots\dots (4)$$

where

\bar{a}_1, \bar{a}_2 : mean of discrimination from both test packages.

\bar{b}_1, \bar{b}_2 : mean of level difficulty from both test packages.

After α and β are known, the result of item parameter estimation and ability parameter from the first test is equated in the similar scale to the second test by using the equation as follow:

$$b_2^* = \alpha b_1 + \beta \dots\dots\dots (5)$$

$$a_2^* = \frac{a_1}{\alpha} \dots\dots\dots (6)$$

$$\theta_2^* = \alpha \theta_1 + \beta \dots\dots\dots (7)$$

where

b_2^* : item difficulty in first test package after equated to second test packages.

a_2^* : item discrimination from the first test packages after equated to second test packages.

θ_2^* : students' ability in first test packages after equated to second test packages.

The accuracy from equating's result can be seen by comparing the mean of Root Mean Square Difference (RMSD) from the students' ability before and after equating process. Kim & Cohen (1996; 17) stated that to count the equating accuracy can use the equation below:

$$\text{RMSD}(a) = \sqrt{\frac{\sum_{i=1}^N (a_2^* - a_1)^2}{N}} \dots\dots\dots (8)$$

$$\text{RMSD}(b) = \sqrt{\frac{\sum_{i=1}^N (b_2^* - b_1)^2}{N}} \dots\dots\dots (9)$$

$$\text{RMSD}(\theta) = \sqrt{\frac{\sum_{i=1}^N (\theta_2^* - \theta_1)^2}{N}} \dots\dots\dots (10)$$

RMSD = Root Mean Square Different

a_2^* = discrimination first test after equated to second test.

a_1 = discrimination first test.

b_2^* = level difficulty first test after equated to second test.

b_1 = level difficulty first test.

θ_2^* = students' ability obtained the first test after equated to second test.

θ_1 = students' ability first test.

III. RESULT AND DISCUSSION

A. Result

1. Equivalent Testing of Test Packages

The proof of the equivalency from both tryout test packages need to be conducted in order to find the parallelism from both tryout test packages. If the both tryout test packages proof as un-parallel, both tryout test packages need to through the equating process. Equivalency testing of both test packages can be carried out with mean different test.

Tabel 1. The Mean Score of Test Package A and Test Package B

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
A	325	6,00	36,00	20,4369	7,96636
B	325	9,00	39,00	23,5292	7,04636
Valid N (listwise)	325				

The mean result of package A tryout test of vocational accounting test is 20.4369 while the mean result of package B is 23.5292. The mean's result from both test packages show that there is the difference where the package B obtained the higher mean's result than the package A.

2. Equating Process

After the tryout of accounting vocational test from both packages prove un-parallel, the equating process for both test packages must be conducted. The equivalence test needs to determine which package used as a benchmark. In this research, the benchmark is package B which means that the package A is equated to package B.

Parameter estimation is based on the students' response suitability to the logistic model. According to analysis by Bilog 3.0 program, it is known that the total item classifies as good and fit most to 2PL. The summary of item parameter from both test package A and package B can be seen as follow.

Tabel 2. The summary of Equating Test Package A and Test Package B

Item Number	Packages A		Packages B	
	Level Difficulty	Item Discrimination	Level Difficulty	Item Discrimination
5	-0,3200	1,3640	0,8430	1,0840
8	-0,6080	1,4440	-0,6770	1,7190
9	-0,1360	1,8420	-0,2820	1,7930
12	-0,3690	1,6340	-0,9490	1,6060
13	-0,4840	1,5480	-0,3070	1,7090
16	-0,2620	1,1970	-0,1540	1,1670
17	-0,7430	1,5080	0,9270	0,6280
20	0,2970	1,1990	0,0660	1,1650
21	1,5110	0,5730	1,5290	0,6300
23	-0,1030	1,5520	0,1690	1,7870
24	-1,1160	0,6060	-0,2700	1,8480
27	-0,0350	1,5910	0,2290	1,6190
30	0,6240	0,9810	0,6820	1,0490
33	0,6020	1,2720	0,7380	0,9100
34	0,4270	1,6010	0,7910	1,4020
36	-0,4680	1,3170	0,3910	0,8870
37	-0,7300	0,6110	-1,3750	0,9050
μ	-0,1125	1,2847	0,1383	1,2887
σ	0,6413	0,3855	0,7490	0,4222

Pursuant to calculation result by Mean/Mean method, it is obtained the constant equation α and β as follow:

$$\alpha = \frac{\bar{a}_1}{\bar{a}_2} = \frac{1,2847}{1,2887} = 0,997$$

$$\beta = b_2 - \alpha b_1 = 0,1383 - (0,9968961 \times -0,1125) = 0,250$$

From the constants α and β above, it is obtained the equating from package A to package B, that is,

$$\begin{aligned}\theta_2^* &= 0,997\theta_1 - 0,250 \\ b_2^* &= 0,997b_1 - 0,250 \\ a_2^* &= \frac{a_1}{0,997}\end{aligned}$$

The conversion equation above shows that there is conversion ability of the students who obtained the test package A to test package B, and also the conversion of item parameter from test package A to test package B. The constants α and β are used in order to make the item parameter transformation so the item parameter is equal.

Tabel 3. The Conversion of Level Difficulty and Item Discrimination from Both Packages by Mean/Mean Method

Item Number	Packages A		Packages B	
	Early Difficulty Level	Early Discrimination	Difficulty Level after Equating	Discrimination after Equating
5	-0,3200	1,3640	-0,0690	1,3681
8	-0,6080	1,4440	-0,3562	1,4483
9	-0,1360	1,8420	0,1144	1,8475
12	-0,3690	1,6340	-0,1179	1,6389
13	-0,4840	1,5480	-0,2325	1,5527
16	-0,2620	1,1970	-0,0112	1,2006
17	-0,7430	1,5080	-0,4908	1,5125
20	0,2970	1,1990	0,5461	1,2026
21	1,5110	0,5730	1,7565	0,5747
23	-0,1030	1,5520	0,1473	1,5567
24	-1,1160	0,6060	-0,8627	0,6078
27	-0,0350	1,5910	0,2151	1,5958
30	0,6240	0,9810	0,8721	0,9840
33	0,6020	1,2720	0,8502	1,2758
34	0,4270	1,6010	0,6757	1,6058
36	-0,4680	1,3170	-0,2166	1,3210
37	-0,7300	0,6110	-0,4778	0,6128
μ	-0,1125	1,2847	0,1378	1,2886
Σ	0,6413	0,3855	0,6394	0,3867

Table 3. explains the conversion from the equating result of discrimination parameter and level difficulty parameter from both test packages. Test package A shows that 17 items test with the mean of level difficulty -0.1125 and standard deviation 0.6413; after equating process, the mean becomes 0.1378 and standard deviation 0.6394. Discrimination parameter of package A has the mean 1.2847 and standard deviation 0,3855; after equating process, it becomes 1.2886 for the mean and 0.3867 for standard deviation. That explanation also shows that item number 5 in package A equals to item number 5 in package B; its level difficulty is -0.0690 and its discrimination is 1.3681. Item number 8 in package A equals to item number 8 in package B; its level difficulty is -0.3562 and its discrimination is 1.4483, etc.

Besides the item parameter, by using the constants α and β , the students' ability who obtained the test package A can be converted to the test package B. Table 4. shows the students' ability who obtained the test package A and the test package B.

Tabel 4. The Conversion of Students' Ability in Both Packages by Using Mean/Mean Method

	Early student's ability obtained test packages A	Students' ability after equating to packages B (θ_2^*)
μ	0,0243	0,2742
σ	0,9823	0,9794

According to Table 4. above, it can be seen that students' ability who obtained the test package A gain the mean 0.0243 and the standard deviation 0.9823. After equating process to package B, students' ability has the mean 0.2742 and the standard deviation 0.9794. It can be

concluded that students' ability who obtained the package A increase the mean and standard deviation after passing through the equating process.

The accuracy of equating test's result by Mean/Mean method is examined by Root Mean Square Different (RMSD) from students; ability before equating and after equating. The calculation is showed as follows,

$$\begin{aligned}\text{RMSD}(b) &= \sqrt{\frac{\sum_{i=1}^N (b_2^* - b_1)^2}{N}} = \sqrt{\frac{1,065}{17}} = 0,250 \\ \text{RMSD}(a) &= \sqrt{\frac{\sum_{i=1}^N (a_2^* - a_1)^2}{N}} = \sqrt{\frac{0,000275}{17}} = 0,0040 \\ \text{RMSD}(\theta) &= \sqrt{\frac{\sum_{i=1}^N (\theta_1^* - \theta_1)^2}{N}} = \sqrt{\frac{20,3035}{325}} = 0,250\end{aligned}$$

The calculation of RMSD in Mean/Mean method shows that the score of RMSD in equating b parameter is 0.250, while RMSD in equating the a parameter is 0.004. The RMSD of the ability shows the score of 0.250. The smaller the RMSD score, the more accurate the equating result.

B. Discussion

The test which was used in the tryout accountancy vocational theory test in *Sleman*, contained of two test packages. This was done as a step to minimize the cheating. Both of two test packages were arranged by *MGMP*, using the similar blueprint but different items. The existence of two test packages results the equivalence of the test is unknown. That could be happen that one of the test package is either harder or easier, *vice versa*. If the test packages were un-parallel, the students' ability could not be seen properly.

The result of mean different score between both test packages shows that there is a differences between them so the equating process must be conducted. The result of the equating process using the Mean/Mean method shows the equation $b_2^* = 0,997bx - 0,250$. The RMSD score is used to see the equating accuracy. The RMSD for parameter b is 0.251, parameter a is 0.004, and parameter ability is 0.250. The smaller the RMSD, the more accurate the equating result.

IV. CONCLUSION AND SUGGESTION

A. Conclusion

According to the result of the research, it can be concluded that both of the packages of tryout accountancy vocational theory test is un-parallel. The equating process needs the equating constants which in this research is conducted by Mean/Mean method. The result of the equating process results the equation $b_2^* = 0,997bx - 0,250$.

B. Suggestion

1. The *MGMP* should conduct the tryout test and item analysis quantitatively so that the test has the better item characteristic.
2. The Educational Authority is better to conduct the training of constructing item test for teacher.
3. The equity of the test need to be considered, although it has been constructed from one blueprint.

REFERENCES

- [1] Educational Ministerial Decree No. 20 of 2007 about Assessment Standard
- [2] Government Regulation No. 19 of 2005 article 68 about National Education Standard
- [3] Kim, S-H., & Cohen, A. S. (1996). A Comparison of Linking and Concurrent Calibration Under Item Response Theory. New York: American Educational Research Association.
- [4] Kolen, Michael J., & Robert L, B, (2004). Test Equating, Scaling and Linking. Methods and Practice. New York: Springer.
- [5] Law No. 20 of 2003 article 57 clause 1 about National Education Standard
- [6] Sukirno. (2007). *Penyetaraan Tes UAN: Mengapa dan Bagaimana*. UNY: *Cakrawala Pendidikan*, November 2007. Vol XXVI pg. 314.

The Effect of Number of Common Items on the Accuracy of Item Parameter Estimates with Fixed Parameter Calibration Method

Dina Huriaty

Mathematics Education Study Program, STKIP PGRI Banjarmasin

e-mail: dina_rty@yahoo.co.id

Abstract—This study aims to determine the effect of number of common items on the item parameter estimates accuracy in fixed parameter calibration methods on the assigned underlying ability distribution. This research is a simulation that was developed based on the item response theory with a three-parameter logistic model. There are the factors in simulation study: (1) three number of common items, namely 10, 20 and 30; and (2) three methods of calibration OWU-OEM (one prior weights updating and one expectation-maximization cycle), OWU-MEM (one prior weights updating and multiple expectation-maximization cycles), and MWU-MEM (multiple weights updating and multiple expectation-maximization cycles), (3) 500 samples for target groups, and (4) the distribution for target group, normal ability distribution. Base group generated to get the parameters of common-items. Twenty-five replications were generated for each condition. The accuracy of method parameter estimation measured by RMSE and absolute-Bias. The research results showed that on normal distribution, the increase of the number of common-items have effect on the accuracy of item parameter estimates on the three fixed parameter calibration method. The small RMSE that obtained of item parameters estimates are not always followed by a small absolute-biased. OWU-MEM method is the most accurate in estimating the parameters in the model with the number of common-items 30.

Keywords: *common-items, estimation, calibration, fixed-parameter*

I. BACKGROUND

When setting the matter for a test, developers can use the items that have been calibrated and added the new test items. In these circumstances, the problem that arises is how to put a new item parameter or parameters item in the previous test, whether it needs to be placed on a items scale that have been calibrated or on a new scale. How to put the estimation parameter from two separate groups into the same scale, can be done by calculating the estimated parameters for each group and then change the scale by using common-items. This can be done through the calibration process. These items which have been calibrated placed as a common-items on new test.

In study by Kim (2006), fixed calibration parameters are divided into five calibration methods. These method is distinguished by the number of ability prior that updated during EM cycle and number of EM cycles is used. These five methods are (1) NWU-OEM (*no prior weights updating and one expectation-maximization cycle*), (2) NWU-MEM (*no prior weights updating and multiple expectation-maximization cycles*), (3) OWU-OEM (*one prior weights updating and one expectation-maximization cycle*), (4) OWU-MEM (*one prior weights updating and multiple expectation-maximization cycles*), dan (5) MWU-MEM (*multiple weights updating and multiple expectation-maximization cycles*). According to Kim, MWU-MEM method is a accurate method in the third set distribution capabilities. On normal ability distribution $N(0,1)$, NWU-MEM and OWU-MEM showed accurate performance on estimating parameter.

Dina & Mardapi (2014) study showed that of the five fixed parameter calibration methods, OWU-OEM is the most accurate method to estimate the parameters of the test item on the national exam in Mathematics.

Based on the results of these two studies and the factors considered in the study of this simulation, this study will only be compared three methods of fixed parameters calibration, namely OWU-OEM (*one prior weights updating and one expectation-maximization cycle*), OWU-

MEM (one prior weights updating and multiple expectation-maximization cycles), dan MWU-MEM (multiple weights updating and multiple expectation-maximization cycles).

Estimate parameters of the common items, then placing it as a fixed parameter, items calibration and then comparing the level of accuracy in estimating the parameters are the background in this study. This study aimed to determine the effect of number of common items on the item parameter estimates accuracy in fixed parameter calibration methods on the assigned underlying ability distribution.

The results of this study are expected to contribute to the development of theory in the field of measurement and testing, particularly for the development of item response theory. The results of this study can provide information about fixed parameter calibration process for various applications such as test score equating and items calibration on test development.

II. RESEARCH METHODS

This study is a simulation. The simulation carried out by conditioning as the real situation on the field. A simulation study is conducted to compare the accuracy of several different models in which analytically settlement can not be obtained easily (Harwell et al., 1997). A simulation study of the generation data involves factors that allegedly influence the accuracy of estimation parameter in this study, which is the number of common items on the capability distribution set. The sample size used in this simulation study is 500 samples for the target-group. The sample size is following minimal sample requirements if using three-parameter logistic model.

In this study, logistics model chosen is the three-parameter logistic model. It is based on the fact that the ability tests that use multiple choice format such as the format of the exam is an example of a situation where the three-parameter logistic model is suitable. Three-parameter logistic model is expressed as follows.

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

(Hambleton, et al. ,1991; Embretson & Reise, 2000; Partchev, 2004).

The ranges for item parameters on three parameter logistic model to estimate the item parameter are as follows. Item difficulty index (b) can be a real number, but normally estimate b is in the range of -3.0 and 3.0 (Hulin, Drasgow & Parsons: 1983, p. 35-36), while according to Hambleton, Swaminathan, & Rogers (1991, p. 13) item difficulty parameter index are in the range -2.0 and 2.0. Item discrimination index (a) has a value in the interval 0.30 to 2.0 (Hulin, Drasgow & Parsons: 1983, p. 35-36), while according to Hambleton, Swaminathan, & Rogers (1991, p. 15), item discrimination index on the scale $(-\infty, +\infty)$. Very rare to find item discrimination parameters index are more than +2. So that the item discrimination parameters are in the interval (0, 2). Asimptut restricted on the interval $0 < c < 1$. Estimates of pseudo-guessing index (c) are often close to $1/m$ for items on the multiple choice test, where m is the number of choices (Hulin, Drasgow & Parsons: 1983, p. 35-36). Based on the limitations of the parameter index, this study set limitation for the item difficulty parameter index is $[-3, 3]$, item discrimination parameter index is $[0.3, 2]$ and pseudo- guessing parameter index is ≤ 0.25 .

There are no guidelines on the number of common items on a test. Kolen & Brennan (1995, p. 248) provides a guide to get the accurate results in the dichotomous model of a single test, the necessary the number of common items at least 20% of the overall number of test items. In this simulation study defined common items are 10, 20, and 30 from a total of 40 rounds on each device that simulated or are respectively 25%, 50% and 75% common items. The distribution considered in this simulation is a normal ability distribution $N(0, 1)$.

III. RESEARCH DESIGN SIMULATION

The factors that considered in the fixed parameter calibration simulation consisting of three types of common items are 10, 20, and 30; the ability distribution for target groups, namely the normal ability distribution; and three methods of fixed parameters calibration, are OWU-OEM (one prior weights updating and one expectation-maximization cycle), OWU-MEM (one prior weights updating and multiple expectation-maximization cycles), and MWU-MEM (multiple weights updating and expectation-maximization multiple cycles). Based on these factor, there are

$3 \times 3 = 9$ conditions. Base-group raised to get the common-items parameters which then set as a fixed parameter.

Harwell (1997) considered that the use of multiple replication is the most appropriate way to conduct simulation studies on item response theory, but Cohen, Kane, & Kim (2001) argues that it is not necessary to have a very large number of replication. A simulation study for item response theory applied only a small amount of replication, which is at least 10 times replication. In this study, each condition performed 25 times replication. Generation of response data is done by Wingen 2.

Parameter estimation accuracy of each replication on a simulation study were evaluated using criteria of RMSE (Root Mean Squared Error) (Kaskowitz & de Ayala, 2001; Kirisci, Hsu, & Yu, 2001; Li & Lissitz, 2000) and the absolute-Bias. The smaller the RMSE, the more accurate the methods used compared to other methods. Beside RMSE, absolute-bias (a-bias) is also an indicator in assessing the accuracy of the calibration method. The smaller the absolute-bias indicate more accurate method to estimate the parameters. The actual parameter values on simulation study are obtained from the output of Wingen 2 (WGI and WGE). RMSE mathematically expressed by the following equation.

$$RMSE = \sqrt{\frac{1}{n} \sum_j^n (\hat{X}_{ij} - X_i)^2}$$

Mean square error (MSE) consists of two components, which is a measure of estimator variability (precision) and bias (accuracy). Bias of an estimator θ of the parameter θ is the difference between the expected value of θ and θ or bias = $E(\theta) - \theta$. Estimator which have bias index equal to zero is called an unbiased estimator, or θ is an unbiased estimator of θ if MSE is equal to the variance of θ . Bias can be negative and positive. Negative bias indicates under-estimated parameters and a positive bias shows over-estimated parameters. Bias is mathematically expressed as follows.

$$\text{Bias} = \frac{1}{n} \sum_j^n (\hat{X}_{ij} - X_i)$$

In this study bias value used is the absolute bias index, to be denoted by a-bias, which is expressed as follows.

$$\text{a-bias} = \left| \frac{1}{n} \sum_j^n (\hat{X}_{ij} - X_i) \right|$$

Software that is used as estimation tool in this study is Parscale. Parscale estimate the parameters of the model responses by Marginal Maximum Likelihood (du Toit: 2003, p. 611). Parscale used to calibrate the fixed parameters by the EM algorithm and can updates the prior ability distribution (Taehoon & Petersen, 2009 and Kim, 2006).

IV. PROCEDURE

Fixed parameter calibration procedure begins by setting the item parameters as common items. Common items Parameter are estimated on the base group, then placed as fixed parameters during the calibration process on the target group. Here is the procedure for each method.

- (1) In OWU-OEM method, a single cycle of EM applied once to new item when the initial prior ability has been updated once. Posterior value calculated after one EM cycle. As NWU-OEM method, in this method only common items that are used to calculate conditional likelihood for each testee.
- (2) The OWU-MEM method estimate the parameters in the new items through multiple cycles of EM, which way to update ability distribution once time. The first EM cycle run in a similar step to the NWU-OEM method, and the value of updated prior will used in the second cycle and cultivated it does not change during the next EM cycle. The initial prior ability renewed

one time after the last cycle of EM . As NWU-MEM method , both common items and new items are used to estimate the distribution of ability. The maximum number of EM cycles is set as 2000.

- (3) The MWU-MEM method update prior ability distribution and find the maximum likelihood estimate on the new item parameters iteration. The first of EM cycle is first performed in the same step as in the method of NWU-OEM. The next step is to update the default value. In the second cycle of the EM, both the item and the ability parameter value are estimated together. After the last cycle of EM, the posterior latent value is obtained. MWU-MEM method uses all the information from the response of new items that use to obtain latent ability distribution and estimate the parameters . The maximum number of EM cycles is set as 2000.

V. RESULTS AND DISCUSSION

The RMSE and absolute-Bias on the factors considered in this study are presented in Table 1.

Table 1. The RMSE and absolute-Bias Index

Item Parameter	CI	OWU-OEM		OWU-MEM		MWU-MEM	
		RMSE	a-Bias	RMSE	a-Bias	RMSE	a-Bias
<i>b</i>	10	0.2384	0.0209	0.2476	0.0314	0.2755	0.0155
	20	0.2340	0.0299	0.2402	0.0458	0.2502	0.0372
	30	0.2234	0.0399	0.2184	0.0538	0.2250	0.0469
A	10	0.2329	0.1070	0.2317	0.1192	0.2534	0.1679
	20	0.2328	0.1313	0.2299	0.1399	0.2437	0.1714
	30	0.2316	0.1612	0.2244	0.1610	0.2336	0.1785
C	10	0.1084	0.0528	0.1071	0.0520	0.1049	0.0510
	20	0.0965	0.0426	0.0951	0.0417	0.0946	0.0424
	30	0.0883	0.0399	0.0864	0.0398	0.0868	0.0405

On normal ability distribution $N(0, 1)$, the increase of the number of common-items affect to the decrease value of the *b*-RMSE on the three methods. The smallest RMSE value appears on OWU-MEM on 30 common-items model. This shows that the increase of the number of common items increases the accuracy of the item difficulty parameter estimation on the three fixed parameter calibration methods.

MWU-MEM method has the smallest a-bias of the item difficulty parameter on the 10 common items model; while on 20 and 30 common items model, the OWU-OEM method has smallest a-bias of the item difficulty parameter. The increase number of common items does not affect on absolute bias indeks decrease of each method. This is shown in Figure 1.

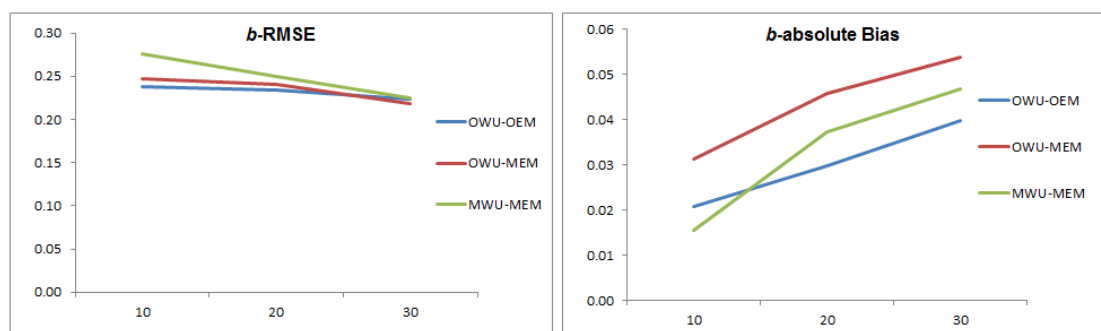


Figure 1. RMSE and a-Bias of Item Difficulty Parameter

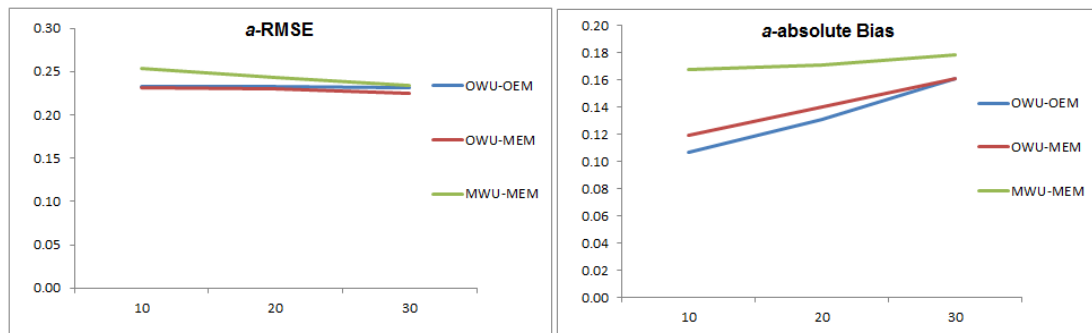


Figure 2. RMSE and a-Bias of Item Discrimination Parameter

The figure 2 shows that on normal ability distribution $N(0.1)$, the increase of the number of common items give an effect to the a-RMSE on OWU-OEM, OWU-MEM, and MWU-MEM method. All methods on sample size of 500 model indicates that if the number of common items increase then a-RMSE will decrease.

As well as a-bias of item difficulty parameters estimates, on a a-bias of item discrimination index shows that the increase the number of common items will increase a-bias indeks. OWU-OEM method has the smallest a-bias index on 10 common items model. OWU-MEM method is the most accurate method to estimate the item discrimination parameters on 30 common items model.

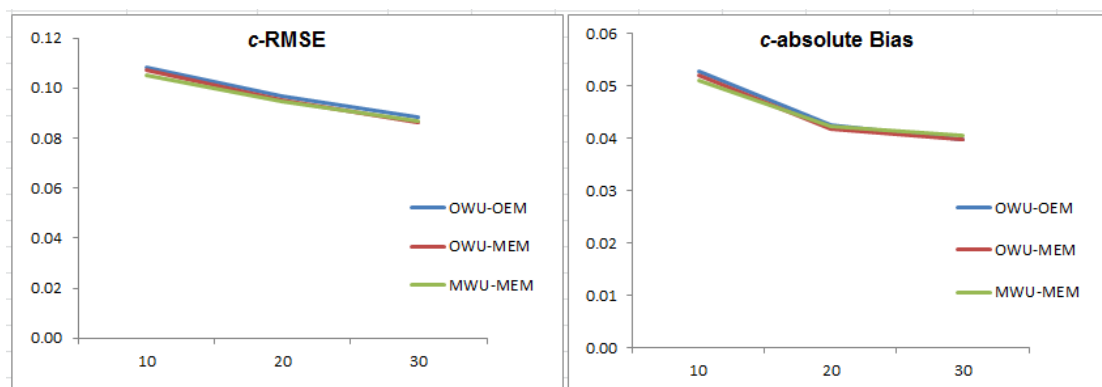


Figure 3. Figure 3. RMSE and a-Bias of Pseudo-guessing Parameter

The figure 3 shows that on the three fixed-parameter calibration methods, c-RMSE is getting smaller as the number of common items increase. OWU-MEM method has the smallest RMSE on 30 common items model. A-bias of pseudo-guessing on all methods are getting smaller as the number of common items increase. MEM OWU method is also the most accurate method to estimate pseudo-guessing parameters based on a-bias index.

This results lead to the conclusion that number of common items can increase the accuracy of item estimation parameter on all fixed parameter calibration methods. These results is relevance with Kim (2006) study, which is the value of RMSE-b and RMSE-a tends decreased when common items increased. According Taehoon & Petersen (2009), when the number of common items increased, each MSE of method deacresed dramatically.

The results of this study showed different results with the results of the study on real data has been conducted by Dina & Mardapi (2014) which the OWU-OEM method is the most accurate method to estimate the parameters. This can be explained that in the real data where only 6 common items, then when analyzed against the accuracy of the method on a model with a common item 10, it appears that the method OWU-OEM shows the smallest RMSE index of item difficulty and a-bias item discrimination. From these results it can be concluded that in models

with a small number of common items, OWU-OEM methods is an accurate method to estimate the parameters.

VI. CONCLUSIONS AND RECOMMENDATIONS

Based on the results of research and discussion, we concluded that the results showed in normal distribution, increasing the number of common-items affect the accuracy of item parameter estimation on three fixed parameter calibration methods. Small RMSE which obtained from item parameter estimation are not always followed by a small a-bias index. OWU-MEM method is the most accurate method to estimate the parameters in 30 common-items.

The implication of this study can be a benchmark in the development of an item bank. The results provide information to developers about how to obtain calibrated test items using a variant of fixed parameter calibration method. The study provides some alternative methods in the process of calibration items and place a number of common items that can improve the accuracy of item estimation parameter which developed for the benefit of the implementation of the test. Those items that have been calibrated can be used to develop an item bank. The existence of item bank can make it easier for the planning and administration the next test.

Limitations in this study are shown in some circumstances, small RMSE values are not followed by a small a-bias, so that the accuracy of the determination of the criteria of a method based on a biased RMSE and smallest by three parameter logistic model can not be met strictly. Prototype results of this simulation study can be a reference in the selection method that can estimate the parameters of the desired items accurately.

BIBLIOGRAPHY

- [1] Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.
- [2] Cohen, A. S., Kane, M. T., & Kim, S.-H. (2001). The precision of simulation study results [Versi elektronik]. *Applied Psychological Measurement*, 25, 136-145.
- [3] Dina Huriaty & Mardapi, D. (2014). Akurasi metode kalibrasi *fixed parameter*: studi pada perangkat ujian nasional mata pelajaran matematika. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18, 188-201.
- [4] du Toit, M. (Ed.) (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.
- [5] Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publisher.
- [6] Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- [7] Harwell, M. R. (1997). Analyzing the result of Monte Carlo studies in item response theory. *Educational and Psychological Measurement*, 57, 266-279.
- [8] Hulin, C.L., Drasgow, F. & Parsons, C.K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.
- [9] Kaskowitz, G. S. & de Ayala, R. J. (2001). The effect of error in item parameter estimates on the test response function method of linking. *Applied Psychological Measurement*, 25, 39-53.
- [10] Kim, S. (2006). A comparative study of IRT fixed parameter calibration methods. *Journal of Educational Measurement*, 43, 355-381.
- [11] Kirisci, L., Tse-chi Hsu & Lifa Yu.(2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality [electronic version]. *Applied Psychological Measurement*, 25, 146-162.
- [12] Kolen, M. J. & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer.
- [13] Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices (2nd ed.)*. New York: Springer.
- [14] Li,-Y. H. & Lissitz, R. W. (2000). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement*, 24, 115-138.
- [15] Partchev, I. (2004). *A visual guide to item response theory*. Downloaded on April 12, 2009.
- [16] Taehoon Kang & Petersen, N. (2009). Linking item parameters to a base scale. *ACT Research Report Series*, 2009-2. Downloaded on September 20, 2010 from <http://www.act.org/research/researchers/reports/pdf/ACTRR2009-2.pdf>.

ANALYSIS OF INTER-RATER CONSISTENCY IN ASSESSMENT FINAL PROJECT FASHION STUDY PROGRAM

Emy Budiastuti

Fashion Study Program (Faculty of Engineering, YSU)

E-mail: emy.budiastuti@yahoo.com

Abstract— This study aims at analyzing score to obtain *Inter-Rater* consistency in the Final Project assessment year 2015 of the Clothing Design Education Study Program based on internal and external assessments. This research is a survey research. The sample of this research is the sixth semester students of Undergraduate Program. The rater as the external experts (academics, practitioners, and association), while the internal experts are the advisors. The coefficient of consistency based on the ICC analysis < 0.50 and Alpha < 0.70 . The ICC criteria is 0.50 at minimum, while Alpha coefficient at least 0.70. The magnitude of ICC coefficient is strongly influenced by the observed value range (variability – variance). Meanwhile, the magnitude of Alpha coefficient is strongly influenced by the number of samples.

Keywords: *consistency analysis, inter-rater, final project*

I. INTRODUCTION

Final Project course is a compulsory course for the students of the Faculty of Engineering of Yogyakarta State University, especially the students of Clothing Design Education study program which they are obliged to take and pass. According to its name, the students are demanded to make a fashion project that is going to display in a fashion show with the theme prescribed. Besides, the students are requested to make a report which later will be presented in the Final Project examination.

There are two stages of the Final Project implementation, i.e. fashion show and oral test. The assessment technique is also adjusted with the two stages. In the fashion show, clothing products made by the students are shown by models, but they must be assessed previously. The assessment of the show is done by the external parties (academics, professional association and practitioners). Meanwhile, the products are assessed by the internal parties, i.e. advisors, and later the results will be combined by the results of oral test.

The assessment of clothing product is started by internal and external parties. They are demanded to give actual scores. *Performance Assessment* of the final project is an assessment that demands the test participant or the students to demonstrate knowledge and skills according to the expected criteria and to be able apply them actually. The students' knowledge and skills can be found out through scoring. Scoring is something that should be concerned, because through scoring, the students' actual abilities will be revealed. The problems which are frequently found in arranging and using skill test lying on validity, reliability and *fairness*.

Performance assessment is an assessment which is done by observing the students' activity while they are doing something. The performance assessment is appropriate to assess competency achievement which demands the students to conduct certain duty, i.e. creating fashion products. This kind of assessment is considered more authentic than written test because the object of assessment reflects the actual ability of the students.

Rater in the Final Project assessment plays important role. *Random error* from the rater will influence the students' score difference comprehensively. There are three sources of mistakes in skill assessment scoring, i.e.: 1) instrument problems; 2) procedural problems; and 3) biased scoring problems. To get result or actual score of the students, the consistency of *inter-rater* is needed much in designing appropriate scoring rubric, selecting and training for *raters*, and *rechecking rater performance*. Based on research conducted by Wainer and Thissen (1993), one of sustainable issues in performance-based assessment is tidak adanya keandalan skala karena

rater. Robert (1981) explained that in order to minimize the measuring mistakes on the performance test, it is suggested to use some raters in assessing the students' behaviors and actually consistency of raters is very determining.

Based on Final Project assessment in the fashion show held previously, the results raise dissatisfaction of the students and advisors. This is because the right of decision is absolutely belongs to the expert. After being observed, there is less understanding or agreement in doing assessment. Therefore, in 2015, the researcher will make efforts to analyze the assessment consistency done by the expert so the dissatisfaction of the students and the advisors will not arise. Through this research, it is expected that there is consistency of *inter-rater* in the assessment of Final Project as expected.

The purpose of this research are: 1) to analyze score in order to create consistency of *Inter-Rater* in the Final Project assessment in 2015 of the Fashion Design Education Study Program based on the external assessment, 2) to analyze score to get *Inter-Rater* consistency in the Final Project year 2015 of the students of Fashion Design Education Study Program based on internal assessment.

The assessment is very important in learning. It must be based on objective, reliable, and trustworthy principles, it should motivate the students to keep on learning. As explained by Kusaeri (2012:16), assessment is a systematic procedure and involved activity of collecting, analyzing, as well as interpreting information that can be used to make conclusion regarding someone's characteristics to determine how extent they achieve learning purposes.

In an assessment there is reliability, referring to assessment consistency which is done by teacher/lecturer. Kusaeri (2012:62-63) states consistency of a measuring activity is how the test score is consistent with measurement one to another. Based on the definition of reliability, some characteristics of reliability are: 1) reliability refers to the results which are obtained through a test instrument, not refers to the instrument itself; 2) reliability is a necessary requirement, but it is not sufficient to fulfill the requirement of validity. A test that is inconsistent in the result will not produce valid information related to the measured ability, so that low reliability can limit the level of validity which is obtained. Reliability produces a consistent result so that the validity will be fulfilled; 3) Reliability is related to statistic, logic analysis. To find out *inter – rater* consistency, the test must be conducted first. Consistency is usually stated in the form of reliability coefficient and *standard error of measurement*.

Assessment is an activity of interpreting the result of measurement, for example succeeded or fail, either good or bad, competent or incompetent, pass and fail, and the kind based on certain standard. Performance test in vocational education usually uses two approaches, i.e. holistic method and analytical method. Holistic method is used when the rater only gives one score (*single rating*), based on the whole assessment from the results of students' performance, while analytical method is used if the rater gives scores on various aspects related to the assessed performance. Analytical rubric is more detailed and it contains statements indicating the measured part or aspect. (Johnson, 2009: 119).

Performance assessment, including clothing making assessment uses assessment criteria, aiming at resulting understanding. The assessment referring to criteria is usually known as reference assessment. According to Djemari (2004:13), the main characteristics that mark the use of reference assessment is score interpretation from the measurement instrument that can make a description regarding ability or knowledge which is owned by learners. The interpretation of the test result is always compared to the standards or criteria which are determined previously.

The Final Project Course is a course that is obliged to be taken and passed for the students of Clothing Design as the final assessment in D3 Program, and as a subject for Undergraduate Program. There are some characteristics of the Final Project course, i.e. 1) as a plus practice course, because after a fashion product finished, it is shown in a fashion show, 2) the making process is started by creating design. In this course, the students are demanded to be more creative either in making design, making pattern, changing pattern, selecting materials, exploring materials, sewing technology, until appropriate to be shown in a fashion show.

The purpose of this research are: 1) to analyze score to get *Inter-Rater* consistency in the Final Project assessment year 2015 Clothing Design Education Study Program based on external assessment, 2) to analyze score to get consistency of *Inter-Rater* in the Final Project assessment year 2015 in the Clothing Design Education Study Program based on internal assessment.

II. RESEARCH METHOD

This research is a survey research, i.e. to obtain an accurate decision from the rater or external and internal experts in the fashion show of the Final Project of Clothing Design Education Study Program year 2015. There are five raters or external experts who are from clothing practitioners, clothing academicians, and clothing profession association. Meanwhile, the internal experts are advisors (lecturers of PT Busana) for each aspect consists of three raters. Population of this research is all students of the Clothing Design Education, while the research sample is the students at the sixth semester in Undergraduate Program. The sample technique applied is *Purposive Sampling*, i.e. the sixth semester students who are taking the Final Project course.

The data are collected by using assessment sheets which are obtained through score documents from the *rater* for all students. After that, the collected score is processed into the final score by using the determined formula that is based on the weight of each assessment aspect. The validity of the instrument that is arranged to present the research data needs to be known. It is obtained by asking opinions of the lecturers according to the field in which they work, i.e. design expert and clothing production expert, trimming expert through study program meeting forum. Then, before the external raters (consists of 5 people) assess, there is a briefing to agree the aspects and criteria which will be assessed.

The data analysis technique applied in this research is descriptive analysis technique, i.e. explaining consistency of the final project measurement from external experts by using *Alpha* and ICC (*Intraclass Correlation*). The ICC analysis is aimed to assess the measuring consistency that is given by some observers who measure the same number.

III. RESULTS AND DISCUSSION

After going through a long process for five months, the students finished the fashion products which were going to be presented in the fashion show. However, before the fashion products are shown, they must be assessed. The assessment of the fashion products were conducted to elect the winner by the external experts who were derived from practitioners, clothing professional association, journalism, fashion analyst and education. The selection of board of expert is considered based on skill and experiences in fashion area.

Futhermore, the assessment activity is started by briefing of all external experts, aiming at consolidating perceptions in each aspect that will be measured. It is done to minimize level of subjectivity and to avoid any mistake during the assessment. Therefore, before the assessment activity is done, the form of assessment for each aspect must be understood and examined seriously to get understanding inter-raters. If there are some experts giving not really high score range, it is predicted that inter-raters have consistent or almost the same score.

A. External Assessment Consistency

The assessment process is started by presenting a finished fashion product. The Final Project assessment's consistency is analyzed by using *Intraclass Correlation (ICC)* and *Alpha* techniques. The criteria of ICC coefficient is 0.50 at minimum, while Alpha coefficient at least 0.80. The ICC coefficient number is influenced by the range of the observed value (variability – variance), while the number of Alpha coefficient is influenced by the sample number. The result of Final Project assessment by the external raters is presented as follows.

Table 1. Summary of ICC and Alpha index from External Raters

No	Study Program	The Assessed Aspects						Inter-Rater Consistency Index	
		Originality		Look		Design Appropriateness			
		ICC	Alpha	ICC	Alpha	ICC	Alpha	ICC	Alpha
1.	PT. Busana (A)	0.31	0.69	0.26	0.63	0.30	0.68	0.29	0.67
2.	PT Busana (D)	0.10	0.34	0.18	0.53	0.20	0.55	0.16	0.47

Based on the results of inter-rater consistency analysis from the external raters, it is shown that the coefficient of consistency based on ICC analysis < 0.50. The low ICC coefficient, there are some aspects which are assessed, i.e. originality, look, and design appropriateness. It is also shown for the coefficient of consistency based on Alpha analysis < 0.70. Thus, the consistency of the assessment for the three aspects show low consistency.

B. Internal Assessment Consistency

Internally, the assessment is conducted by the advisors. The assessed aspects are: trimming, clothing technology, design appropriateness, and design. Below is presented inter-rater consistency index, both using ICC or Alpha.

Table 2. Summary ICC and Alpha index from Internal Raters

No	Study Program	The Assessed Aspects								Average of <i>Inter-rater Consistency Index</i>	
		Trimming		Clothing Design		Wearing Appropriateness		Design			
		ICC	Alpha	ICC	Alpha	ICC	Alpha	ICC	Alpha	ICC	Alpha
1.	PT. Busana (A)	0.26	0.51	0.62	0.83	0.57	0.82	0.67	0.86	0.53	0.76
2.	PT Busana (D)	0.45	0.71	0.62	0.83	0.25	0.50	0.42	0.68	0.44	0.68

The results of *inter-rater* analysis consistency from the internal raters (advisors) show that the coefficient of consistency based on ICC analysis < 0.50 is only found in trimming assessment consistency. The aspects of clothing technology, wearing appropriateness and design have fulfilled the required ICC coefficient that is > 0.50 . The Alpha coefficient shows that trimming aspect and clothing technology aspect show good consistency. Meanwhile, for other aspects which are wearing appropriateness and design show coefficient < 0.7 . It is concluded that the inter-rater consistency is low.

Based on the results of consistency based on *Intraclass Correlation (ICC) and Alpha*, it shows that external inter-raters do not show high level of understanding or consistency or similarity in scoring the results of the students' fashion products. After doing deep observation, it is found that there is incomplete instrument, such as rubric. Although there are other instruments, without clear rubric it will make the assessment in determining score unclear. If the assessment instrument is available before the assessment is conducted, it will make the raters easier in determining scores. The number of the inter-rater scores which are almost the same will decide the actual students' competency. Thus, in determining the winner, there are scores which are almost the same or consistent. If the students get similar scores or almost the same, it will be accepted sincerely without disappointed, mad, sad and unfair feelings and so on toward the final products.

IV. CONCLUSION AND SUGGESTION

It can be concluded that: 1. Based on the results of inter-rater consistency analysis from external raters, it is shown that the coefficient of consistency is based on ICC analysis < 0.50 , while the coefficient of consistency based on Alpha analysis < 0.80 . Based on the results of inter-rater consistency analysis from the internal raters (advisors), it shows that the coefficient of consistency based on ICC analysis < 0.50 , while the coefficient of consistency based on Alpha analysis < 0.80 . However, the interrater consistency of each aspect shows good consistency.

Some important suggestions can be concluded are: 1) regarding the low consistency coefficient of the external experts, either by using ICC coefficient or Alpha coefficient, for the next fashion show, it needs to be selective in determining experts, to get decision regarding the students' works accurately; 2) in order to get good consistency of the next Final Project, complete and clear assessment instruments are needed.

REFERENCES

- [1] M Djemari, "Pengembangan sistem penilaian berbasis kompetensi," in *Proceeding: Rekayasa sistem penilaian dalam rangka meningkatkan kualitas pendidikan*, Yogyakarta: HEPI, 2004.
- [2] Kusaeri, Suprananto, "Pengukuran dan penilaian pendidikan," Yogyakarta: Graha Ilmu, 2012.
- [3] R.L. Johnson, J.A. Penny, B. Gordon, "Assessing performance: designing, scoring, and validating performance task," London: The Guilford Press, 2009.
- [4] S. A. Yorkovich, G.S. Waddell, R.K. Gerwig, "Competency-based assessment systems: Encouragement toward a more holistic approach," http://spiritoorganization.com/documents/Waddell_CompetencyBasedAssessment.pdf, 2008. (Date of access: 5 Januari 2010)

USING FUZZY LOGIC TO SELECT ITEM TEST IN COMPUTERIZED BASE TESTING

Haryanto

haryanto.ftuny@gmail.com

ABSTRACT --- The development of the computerized base testing (CBT) with fuzzy logic algorithm aimed to investigate: the performance of fuzzy inference system in making decision to select the appropriate test items for the students. This research and development (R&D) study consists of two parts: (1) developing CBT program with *fuzzy* logic, (2) the CBT program testing was conducted to the samples from the students at Senior High School of Yogyakarta. Data were collected through observation, documentation, questionnaire and test. The data were analyzed using quantitative descriptive technique. The result of research showed that: (1) the CBT program is able to work properly according to the function and the type of users, and (2) the CBT program can managing (a) administer the test *bank*, (b) organize test items automatically, (c) organize the test items based on students' competence, (d) randomize the answer in the options, and (e) record the result of the test simultaneously and individually. Based on the data analysis the result of applying the CBT program for testing students' competence shows that: (1) the inference system of the program could appropriately provide adapted test items to the students based on the correct-incorrect responses they gave, (2) the collection of test items provided by the program to each participant can appropriately describe the competence of each students. In conclusion, the CBT program is using *fuzzy* logic algorithm can perform its function well to select the suitable test items and to measure the result of students' achievement in the learning process.

Keywords: *Fuzzy logic, Inference systems, Item tes, Adaptive testing*

I. INTRODUCTION

The development and advancement of Science, Technology and the Arts (Science and Technology), particularly the field of Information Technology (IT), showed a significant increase for the benefit of education. It can be seen, among others: the use of IT for learning the subjects in the curriculum, the learning process with media-based IT, process test computer-assisted (PP No: 22, 2005; Baumgartner & Jackson, 1995; Bates & Poole, 2003; Chee & Wong, 2003).

Target of increasing the quality of education, through the measurement of IT-based, has sought to revive (Chee & Wong, 2003). Use of IT for testing purposes intended for effectiveness and efficiency of the implementation and operation of the test (Chee and Wong 2003 and Towndrow & Vallence 2004). IT era through Computerized Base Test (CBT), the computer is programmed to process the order to see the grains exam questions and receive answers from the participant's response (Chee and Wong, 2003; Towndrow & Vallence. 2004).

Based on observations of the programs used by institutions / agencies organizing computer-based testing, there are some things that need to be studied and researched further. The process of making random test item bank of questions in computer-based testing, raises the problem of mismatch between the level of difficulty about the ability of the test taker. On the other hand, retrieval of test items sequentially question bank that is widely used, weakens the validity of the test results.

Recent developments, along with the bank about the business of making and improving the quality of tests, computer technology is very helpful for this purpose (Hambleton, Swaminathan, & Rogers, 1991). Improved quality conformance test item is intended to test the ability of the test taker or often referred to as "Computerized Adaptive Test / CAT)". Called computerized (computerized) for the implementation of testing using the computer. Adaptive, that the level of difficulty of test items that appear based on the results of the analysis algorithms in a program, to match the abilities of participants. Suitability items are based on the responses of right and wrong answers to the test taker.

Various theories of evolution of computing to support human needs are constantly sought. One of these developments for the benefit of the test is based computational algorithms "Artificial Intelligence (AI)" or artificial intelligence (Luger, 2005). This theory describes an attempt to make the engine computer can "learn" so that it can work with the "intelligence" as humans. Problems associated with testing the learning outcomes, the computer is programmed with the provision of knowledge to determine the ability of test takers through intelligent analysis of the responses to participants' answers to the test, to choose the appropriate test items.

AI approach in this study is fuzzy logic. Fuzzy logic inference systems are used for the purposes of (decision) to a problem that is both qualitative and quantitative (Yan, Ryan, & Power, 1994). Application of the test is to select test items are appropriate and in accordance with the ability of the test participants, in order to obtain the results in the form of a valid measurement.

Based on the description above the expected research goals are to generate:

1. Model CBT program inference system with fuzzy logic algorithm to select test items right.
2. CBT program that is capable of providing reliable test results informai.

II. CBT MODEL WITH FUZZY LOGIC

Hambleton, Swaminathan, & Rogers (1991) in his book "Fundamentals of Item Response Theory" says that what is meant by testing adaptive computerized constitute "would be to give every examinee a test that is" tailored "or adapted, to the examinee's ability level", It also said that tests with long items have been fixed, it is not efficient if given to all participants of the test. This was due to the ability of diverse participants should reasonably be given only to a few questions that correspond to the level. He said that the test can be shortened without losing precision test, if the test items prepared (administered) in accordance with the capabilities of each participant tests. For that, of course, requires complete information about the ability of the test participants, so that each participant will receive a set of test questions according to his ability (administered a unique set of items).

Masters & Keeves (1999) mentions that in the adaptive test, test items were selected from a question bank, based on the rule (rule) the selection of test items previously defined. Programmable computer with the ability to have a high convenience for use in the testing process is adaptive (adaptive testing).

Baeck, Fogel, & Michalewicz, (1997), the outline suggests that the evolution of computing in the field of measuring instruments and measurement starts from a conventional classical, to modern intelligent. According to Goldberg (1989) and Mitchell (1997), the desired goal improvement (evolution) computing is to make the engine computer not only as a tool to count, but also optimizing the workings of the machine that has the ability and intelligence of an expert (teaching, scoring, testing, and interpreting) humane (humane = able to learn the character of testee's ability to be able to judge correctly). Understanding adaptive in this research is the ability to customize the item difficulty with the test participants' ability test (the examinee's ability) based on the response of right and wrong answers. The ability to computer hardware programmed with fuzzy logic.

Model testing is modern with IRT differentiated by the number of parameters of test items, namely the model of the parameter (Rasch model), two parameters, and the three parameters (Hambleton, Swaminathan, & Rogers, 1991). Van der Linden & Hambleton (1997), mentions these parameters are difficult items, different power items, and guesses. IRT models for dichotomous two-parameter test items (item difficulty, item different power) is as follows (van der Linden and Hambleton (1997); Hambleton, Swaminathan, & Rogers (1991); Lord (1980)):

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad \text{atau} \quad P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad i = \text{test item } 1, 2, 3, \dots, n$$

$P_i(\theta)$: probabilitas peserta dengan kemampuan θ untuk menjawab butir ke- i dengan benar.
 θ : tingkat kemampuan peserta uji
 n : banyaknya butir tes
 e : nilai transcendental yang besarnya 2,718
 b_i : tingkat kesulitan butir
 a_i : daya beda butir

Figure 1 shows the characteristic curve and distribution capabilities of the two groups, namely the poor performance and high capability. It is seen that the group with high ability to have a greater probability of correctly answered the item correctly than the group with low ability. Thus the characteristics of the item does not depend on the group.

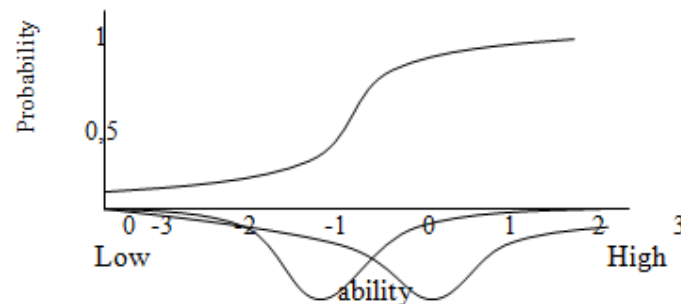


Figure 1. Item Characteristic Curves and Distribution Capabilities
(Adopted from Hambleton, Swaminathan and Rogers (1991))

Rasch model assumptions are based on the specific requirements of objectivity (Hambleton, Swaminathan, & Rogers, 1991). Key assumptions: (1) their unique ordering (unidimensionality) according to people's ability and item difficulty of the test. (2) the implications of the assumptions to-1, the local independence, the value of the likelihood of people answering correctly a test item does not depend on the answers to other test items. (3) are also implications of the assumptions to-1, namely equality of discrimination, interference ratio which is represented by the slope (slope) maximum grain characteristic curve assumed to be similar for all test items. (4) unidimensionality set of test items in a test device measures only one dimension. (5) conduct random guessing (random guessing behavior). Rasch Model requires that for each test item, the value of the possible answers correct asymptotik tend to zero in line with the declining level of ability. (6) when the ability of people increases, the value of the possibility of correct answers to a test approaching one (as opposed to assuming all four).

Fuzzy logic system is one branch of science that studies on artificial intelligence (artificial intelligence / AI). The basic principles of intelligent systems (Luger, (2005); Nilsson, (1980)) is to make the computer through certain programming techniques to be able to think, take the right decisions and act in ways as humans do.

The ability of fuzzy logic in describing the student's ability to use reasoning monotonous. Through the monotonous reasoning, will obtain a degree of certainty regarding the ability of students is based on responses provided during the testing process. The certainty factor is a unity of fuzzy logic in order to describe the ability of students. If there are two fuzzy areas related by simple implications, namely:

IF x is A, THEN y is B.

The transfer function is expressed by the notation $y = f((x, A), B)$.

Mathematically described as follows:

$$\mu_A[x] = (x-a)/(b-a) = q \quad \text{and} \quad \mu_B[y] = 1-2[(d-y)/(d-c)]^2 = q$$

In this case the known value of x to find the value of y through the value of q as reasoning monotonous (Figure 2 shows a graph of the process).

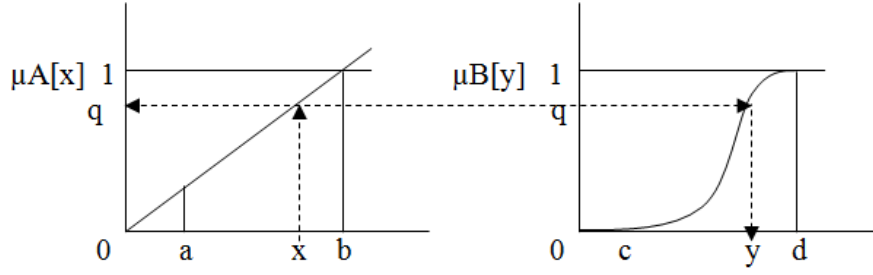


Figure 2. Reasoning Monotone

The value of the output of fuzzy inference system in the form of item difficulty is further given to test participants. Test items taken from the banks about the difficulty level according to the results of fuzzy logic inference. During the exam, the test participants' ability calculated estimated using maximum likelihood. Ability obtained in the form of a score of -3 to +3. In its application, the result of the ability to be converted into a value of 0 to 10 or 0 to 100.

An assessment of the ability of students, first performed by calculating the value of $p_i(\theta)$ and $q_i(\theta)$ of each item test. In this study, using two parameters, namely: the power difference (a_i) and the index of difficulty (b_i), thereby calculating the value of $p_i(\theta)$ (chance-takers with the ability of $[\theta]$ answered grain test all i really) and $q_i(\theta)$ (odds-takers with the ability $[\theta]$ responded to the test item- i wrong) is:

$$p_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}} \quad p_i(\theta) + q_i(\theta) = 1$$

Furthermore, with the known value of $p_i(\theta)$, $p_i(\theta)$ and θ can be calculated likelihood values $L(U | \theta)$ with the formula:

$$L(U | \theta) = \prod_{i=1}^n p_i^u q_i^{1-u}$$

where:

n : a lot of test items

u : the students' answers on the test item

To determine the ability of test takers, calculated beforehand using Maximum Likelihood Estimation $L(\theta | U)$ with the formula:

$$L(\theta | U) = \frac{L(U | \theta)}{\sum L(U | \theta)}$$

Based on the results of $L(\theta | U)$ to the value of θ from -3.0 to 3.0, the estimate of the student's ability test participants is the value of θ from the $L(\theta | U)$ is the highest (maximum). So estimation capability test participants determined by the formula:

$$\text{Estimate } \theta = \text{Maximum } [L(\theta | U)]$$

Such estimates implies that the participants opportunities to test the ability of $[\theta]$ answer test items $\max L(\theta | U) \times 100\%$ correct. On the other hand, means the chances of the test participants with the ability of $[\theta]$ answer test items $[1 - \max L(\theta | U)] \times 100\%$ wrong.

III. RESEARCH METHODE

This study uses research approaches Research and Development. There are two stages in the implementation process, namely: the development phase of CBT with fuzzy logic algorithm and the implementation phase of the product. In the product development stage, the process is done is referring to the model Pressman (1997) and Rolston (1988). The second stage is to apply the product by following the steps proposed by Borg and Gall (1983).

The study was conducted in: (a) Laboratory of Computer Education Electrical Engineering FT UNY, for process development / product development CBT program. (B) SMA Negeri 6 Yogyakarta Special Region of Yogyakarta, as a place for the implementation / research products CBT program.

Subject of the study include: (a) Teachers of Mathematics, to the process of extracting information and identification requirements necessary systems in CBT program. (B) Students to test the functionality and performance of products CBT program.

Data required in this study include: (a) The qualitative data in the form of the required functions of the program, potential users of the program, and minimum standard hardware used program, and the results of internal testing of the functionality of the program. (B) The quantitative data in the form of teacher responses to product CBT program and the results of performance testing program in the process of student learning outcomes.

Data collection techniques include: (a) For the development of the product, using: (1) observation, concerning: the accuracy of instruction, verification and validation of products (alpha-beta testing). (2) The questionnaire and interview, regarding: the completeness and accuracy of the program's functionality. (3) Documentation, regarding: materials, shapes and models tests. (B) For the implementation of the product, using: (1) observation, regarding: truth, accuracy, functionality, and product capabilities in analyzing the ability of students. (2) Documentation, concerning: the data value student achievement test results.

Data analysis technique used is quantitative descriptive and evaluative. Quantitative descriptive, carried out to assess the fuzzy inference system in selecting test items corresponding to the level of ability of the test taker. Evaluative descriptive analysis technique conducted to determine the feasibility, capability and effectiveness of CBT in its function of measuring the ability of the test taker.

IV. RESULT RESEARCH

Ratings response to CBT program, conducted through questionnaires to teachers of mathematics studies. Data is collected after the CBT program teachers use for various activities in accordance with their authority.

Scores of teacher responses to the CBT program at each aspect in the range of > 3 to 4. The teacher's response to CBT programs on aspects of performance (1) Operational / use program, (2) the program display (3) the relevance of the program with the test material (4) the benefit of the program in support of teachers' duties, obtaining very good value. It shows that the response of teachers to CBT program with fuzzy logic as a whole is considered very good, or worth using.

Beta testing is used to determine the ability of CBT program work in measuring the ability of students through the test items done. Grain test question bank which is equipped with a difficulty index parameters and different power. Based on the test results can be described a number of the group's ability as follows: (1) There were 17 students expressed less with categories ability -3 to <1. (2) There are 7 students expressed both by category ability ≥ 1 to <2. (3) There were 14 students expressed very well by category ability ≥ 2 to 3.

Description accuracy CAT programs work on the selection of test items given to students illustrated by the analysis of the average correlation values and standard deviations in theory with the work of fuzzy inference CBT program. Results of correlation analysis of the level of difficulty of test items highest correlation values (r) of 0.93 and the lowest (r) of 0.09 and an average value of correlation (r) of 0.72. It shows that CBT program with fuzzy logic inference systems managed by either selecting test items that match the students' abilities. That is, that the test items were selected from a question bank between the theoretical analysis with the work of the CBT program has a high relationship (mean $r = 0.72$).

The results of the analysis of standard deviations (sb) is theoretically about the difficulty level of test items obtained the highest score of 0.12 and the lowest $sb = sb = 0.03$, and the average value $sb = 0.09$. Sb analysis results are fuzzy about the level of difficulty of test items obtained the highest score of 0.15 and the lowest $sb = sb = 0.05$, and the average value $sb = 0.11$. It also shows that CBT program with fuzzy logic managed by either selecting test items that match the ability of the test taker.

MLE analysis results were also obtained Likelihood function graphs concerning the description of the participants' ability tests. The ability to graph functions obtained from the

analysis of the level of difficulty grains taken from a question bank. Values illustrated by the ability of the curve with a limit of $-3 < \theta < +3$ it shows that the ability of students in normal distribution. Based on the analysis of data as a whole, shows that the fuzzy logic inference systems managed by either selecting test items are appropriate and in accordance with the ability of the test taker. In that case, the test taker with a high capacity majority receiving test items with a high difficulty level. Participants test-skilled receiving test items with a moderate level of difficulty. Participants test with low ability majority receiving test items with a low degree of difficulty. On the other hand, a lot of grain received test takers with one another are also diverse (not the same). Participants test with high ability and low ability or receive many grains of less than capable of being. It shows that CBT program with fuzzy logic has also been successful with a nice set many items are given according to student ability (adaptive).

CBT program developed can also display packaging test items to the location of randomized answer options. This meant that the inter-takers can not cooperate with each other in answering the test. The uniqueness and at the same time more value from product CBT program is the use of artificial intelligence techniques (artificial intelligence) for the development of fuzzy logic inference systems. CBT program product is also capable of: packing the test items automatically, carry out the administration of test items, save the test results, and do the scoring.

The ability of the CAT program with the fuzzy logic algorithm as described above, has been as expected by item response theory. That each student receives the difficulty level test items that match the characteristics of the item information. On the other hand, each student will receive a number of test items that vary according to their abilities. This is in accordance with the demands of the nature of adaptability tests of CAT

V. CONCLUSIONS

The conclusions that can be drawn based on the data analysis and discussion are as follows: (1) Model CAT with fuzzy logic algorithm is able to select test items with the appropriate level of difficulty based on right and wrong answers students. In addition, it is also able to arrange a lot of test items each student according to their ability level. (2) CAT models inference system with fuzzy logic algorithm is able to set the level of difficulty of test items for each student according to their ability level. (3) The accuracy of the system inference in selecting test items that match the students' answers in response average value error / standard deviations are 0.093 and 0.112 in fuzzy theory. Based on the graphical analysis and correlation of test items between theoretical analysis with fuzzy results with the average value of $r = 0.72$ per student. (4) The ability of the model CAT with fuzzy logic algorithm to estimate the students' abilities based test items received is obtained: (a) 34% of students are very good capacity (b) 21% of students the ability to both, and (c) 45% of students less ability good. (5) The findings of this study, are: (a) CAT models inference system with fuzzy logic algorithm is able to conduct elections in an adaptive items. (b) System inference CAT models with fuzzy logic algorithm: Able to manage database question bank for the computerized exam. Being able to choose adaptively test items in accordance with the students' ability in computerized exam.

REFERENCE

- [1] Baeck, T., Fogel, D.B., & Michalewicz, Z. (1997). "Handbook on evolutionary computation". New York: IOP Press.
- [2] Baumgartner, T.A., & Jackson, A.S. (1995). "Measurement for evaluation in physical education and exercise science" (5th ed.). New York: WCB Brown & Benchmark Publishers.
- [3] Bostock, S.J. (1997). "Designing web-based instruction for active learning". Englewood Cliffs: Educational Technology Publications.
- [4] Borg, W.R., & Gall, M.D. (1983). "Educational research: An introduction" (4th ed.). New York: Longman Inc
- [5] Brennan, R.L. (2006). "Educational measurement" (4th ed.). New York: American Council on Education.
- [6] Chee, T.S., & Wong, A.F.L. (2003). "Teaching and learning with technology". Singapore: Prentice Hall.
- [7] Echols, J.M., & Shadily, H. (1986). "Kamus Inggris Indonesia". Jakarta: PT. Gramedia.
- [8] Eiben, A.E., & Michalewicz, Z. (1999). "Evolutionary computation". New York: IOP Press.
- [9] Goldberg, D.E. (1989). "Genetic algorithms in search: Optimization & machine learning". New York: Addison-Wesley.
- [10] Hambleton, R.K., Swaminathan, H., & Rogers, H. (1991). "Fundamentals of item response theory". New Delhi: Sage Pub. Inc.
- [11] Hornby, AS. (1984). "Oxford advanced learner's dictionary of current English". London: Oxford University Press.
- [12] Lin, R.L. (1989). "Educational measurement" (3rd ed.). New York: American Council on Education, Macmillan Publishing Company.

- [13] Lord, F.M. (1980). "Applications of item response theory to practical testing problems". Englewood Cliffs: Lawrence Erlbaum Associates, Publishers.
- [14] Luger, G.F. (2005). "Artificial intelligence, structure and strategies for complex problem solving" (5th ed). New York: Addison Wesley.
- [15] Masters, G.G., & Keeves, J.P. (1999). "Advances in measurement in educational research and assessment". New York: Pergamon Press.
- [16] Mitchell, T.M. (1997). "Machine learning". New York: McGraw-Hill Companies, Inc.
- [17] Nilsson, N.J. (1980). "Principles of artificial intelligence". New York: Tioga Publishing Company.
- [18] Pressman, R.S. (1997). "Software engineering: A practitioner's approach". New York: Mc Graw hill Book, Inc.
- [19] Rolston, D.W. (1988). "Principles of artificial intelligence and expert systems development". Singapore: McGraw Hill Book, Co.
- [20] Terano, T., Asai, K., & Sugeno, M. (1992). "Fuzzy systems theory and its applications". New York: Academic Press, Inc.
- [21] Tim Prima Pena (2007). "Kamus besar bahasa Indonesia". Jakarta: Gita Media Press.
- [22] Towndrow, P.A., & Vallence, M. (2004). "Using IT in the language classroom: A guide for teachers and students in Asia" (3rd ed.). Singapore: Longman Pearson Education South Asia Pte. Ltd.
- [23] van der Linden, W.J., & Hambleton, R.K. (1997). "Handbook of modern item response theory". New York: Springer Verlag.
- [24] Wang, L.X. (1994). "Adaptive fuzzy systems and control: Design and stability analysis". Englewood Cliffs: Prentice Hall Inc.
- [25] Widrow, B., & Stearns, S.D. (1985). "Adaptive signal processing". Englewood Cliffs: Prentice Hall Inc.
- [26] Yan J., Ryan M., & Power J. (1994). "Using fuzzy logic". Englewood Cliffs: Prentice Hall Inc.
- [27] Zimmermann, H.J. (1991). "Fuzzy set theory: And its applications" (2nd ed.). London: Kluwer Academic Publishing.

An Application of the Generalized Logistic Regression Method in Identifying DIF (Analysis of School Examination in Soppeng)

Herwin

Yogyakarta State University

winunm@gmail.com

Abstract - This study aimed to analyze the DIF in school exam test in Soppeng (South Sulawesi). This research is a quantitative research focusing Elementary School Examination of Mathematics with 40 items multiple-choice and examinees answer sheet as much as 400 sheets in which 200 sheets (73 Men and 127 Women) is a response to the examinees of the rural, as well as 200 sheets (70 Men and 130 Women) is a response to the examinees of the urban. Data collected by the documentation, and the data were analyzed with a Generalized Logistic Regression Method. The results showed that there are 6 items detected Uniform DIF (5, 6, 16, 18, 27, and 38). There are two items detected nonuniform DIF (16 and 19), and there are 3 items detected Both Type of DIF (5, 16, and 27).

Keywords: DIF, Generalized Logistic Regression Method

I. INTRODUCTION

Assessment is an important component in education. One effort that can be done to improve the quality of education is to improve the quality of assessment systems. The good assessment system that will encourage teachers to define good teaching strategy in motivating students to learn better. Therefore, in an effort to improve the quality of education needed improvement assessment system applied. From this assessment, teachers will gain the ability portrait or profile of students to achieve basic competencies defined in the curriculum.

Good examination system should be supported by the various components. One of the components supporting the implementation of good exam is the quality assessment instrument. One indication of a good instrument is an instrument to be free from the threat of bias. Bias is defined as a systematic error in the measurement process (Osterlind 1983: 10). The bias is negative conditions that could threaten the quality of the test. Item bias could harm a group of students by gender, religion, ethnic, social, and economic benefits of different and other groups, when in fact the two groups that have the same knowledge and skills.

Differential item functioning (DIF) procedures are currently the dominant psychometric methods for addressing fairness in standardized achievement, aptitude, certification, and licensure testing (Jodoin & Gierl, 2001: 329). DIF occurs when examinees from different groups show differing probabilities of success on (or endorsing) the item after matching on the underlying ability that the item is intended to measure (Zumbo, 1999: 12). The purpose of DIF analyses is to detect response differences of items in questionnaires, rating scales, or tests across different subgroups (e.g., gender), while controlling for ability level (Zhang, 2015: 1). Bias conditions are a threat within a test. Bias can arise as a matter of invalid eg in terms of construction problems or content not suitable for a particular group of students. The existence of the item bias in an instrument could affect the test results were not good. That is, the decision obtained from the test will be less objective if the instrument contains an item bias. The presence of item bias must be detected so that the test execution can be better qualified.

Given the function and purpose of the test execution is very important, the measurement instruments used must be of quality (free from bias). Therefore, detection item bias on measuring instruments is essential. Item bias procedures are used to determine whether the individual items on an examination function in the the same way for two groups of examinees, usually defined by racial and ethnic background, sex, age and experience, or condition of handicap (Scheumnean & Bleistein, 1999: 220).

There are many different ways that can be used in detecting item bias, but in this research the method is Generalized Logistic Regression Method. The generalized logistic regression method (Magis, Raiche, Beland and Gerard, 2011) allows for detecting both uniform and non-uniform differential item functioning among multiple groups without requiring an item response model

approach. It consists in fitting a logistic model with the matching criterion, the group membership and an interaction between both as covariates. The statistical significance of the parameters related to group membership and the group-score interaction is then evaluated by means of the usual likelihood-ratio test.

A. Research Problem

Based on the background of this study, the research problem is "how to charge DIF on the instrument of School Examination in Soppeng based Generalized Logistic Regression Method".

B. Objectives

Based on this research problem, the goal of this research is "to know the charge DIF on the instrument of School Examination in Soopeng based Generalized Logistic Regression Method".

C. Significance of Research

This study is expected to be useful:

1. For the teacher, it can be used as a reference for applying Generalized Logistic Regression Method separately for detecting DIF. In addition, the results of this study can be used as an instrument of information about the characteristics of the tests used in School Examination in Soppeng.
2. For the Department of Education Soppeng, this research useful as a basis in determining the policy for implementation of School Examination in the next year.

II. RESEARCH METHOD

Generally, this study uses a quantitative approach. This study was conducted in an elementary school in Soppeng (South Sulawesi) on the Academic Year 2015/2016. This study focused on Instrument of School Examination (Mathematics) on the Academic Year 2014/2015 totaling 40 items. Test in the form of multiple choice with the option of four options developed by KKG Team appointed by the authorities. Data collection techniques in this study is documentation techniques.

To apply the Generalized Logistic Regression Method, then it takes the response patterns of examinees coming from the answer sheets of the examinees. In this study, the answer sheets of the examinees were sampled at random in order to obtain 400 answer sheets of the examinees in which 200 sheets (73 Men and 127 Women) derived from the responses of examinees from rural and 200 sheets (70 Men and 130 Women) derived from responses examinees from urban areas.

Data were analyzed using Generalized Logistic Regression Method. The generalized logistic regression DIF model, as presented by Magis et al (Svetina, & Rutkowski, 2014: 5), has the following form:

$$\text{logit}(\pi_{ig}) = \alpha + \beta S_i + \alpha_g + \beta_g S_i \quad (1)$$

Where π_{ig} is the probability of examinee i from group g correctly responding to an item, logit is the natural log of the odds of correctly answering an item, α and β are common intercept and slope parameters (i.e., for all groups), α_g and β_g are group-specific slope and intercept parameters, and S_i is the total test score for examinee i , which serves as a matching variable and a proxy for the ability level of the examinee. For model identification purposes, group-specific parameters for the reference group (denoted as $g = 0$), α_0 and β_0 , are set to zero. In other words, if $g = 0$ (i.e., reference group), $\text{logit}(\pi_{ig}) = (\alpha + \alpha_g) + (\beta + \beta_g) S_i$. (2)

According to Magis et al. (2011), three types of DIF can be investigated using this framework: a) uniform DIF (UDIF), b) nonuniform DIF (NUDIF), and c) both types of DIF effects together (UDIF and NUDIF effects simultaneously). Tested null hypotheses for these three types include:

1. $H_0: \alpha_1 = \dots = \alpha_f \mid \beta_1 = \dots = \beta_f = 0$ **UDIF**
2. $H_0: \beta_1 = \dots = \beta_f = 0$ **NUDIF**
3. $H_0: \alpha_1 = \dots = \alpha_f = \beta_1 = \dots = \beta_f = 0$ **both (UDIF & NUDIF)**

To determine the item that contains DIF necessary a criteria. DIF is used based on the criteria of advice Svetina, & Rutkowski (2014: 5) that using maximum likelihood, we can test the null hypotheses in (1) to (3) by likelihood ratio test (LRT) as a criterion to detect DIF, where the null (M_0) and alternative (M_1) models were given by:

$$M_0 \equiv \text{logit}(\pi_{ig}) = \begin{cases} \alpha + \beta S_i + \alpha_g & \text{in NUDIF} \\ \alpha + \beta S_i & \text{in BOTH and UDIF} \end{cases} \quad (3)$$

and

$$M_1 \equiv \text{logit}(\pi_{ig}) = \begin{cases} \alpha + \beta S_i + \alpha_g + \beta_g S_i & \text{in BOTH and NUDIF} \\ \alpha + \beta S_i + \alpha_g & \text{in UDIF} \end{cases} \quad (4)$$

DIF is then tested by the Wilks lambda statistic, where $\Delta = -2\log\left(\frac{L_0}{L_1}\right)$, L_0 and L_1 are the corresponding maximum of the likelihoods for the M_0 and M_1 , respectively, which follows an asymptotic chi-square distribution with degrees of freedom of the asymptotic null distribution of Δ equal to 2 df for BOTH type (UDIF & NUDIF) and df for both the UDIF and NUDIF. Item would be flagged as DIF when the lambda statistic is sufficiently large, or stated alternatively, when the alternative model is preferred ($p \text{ value} \leq \alpha$ or GLR Statistic > Threshold chi-square).

III. RESEARCH FINDING AND DISCUSSIONS

Based on the results of this study are described three types of DIF testing are: 1) UDIF , 2) NUDIF, and 3) Both (UDIF and NUDIF effects simultaneously).

A. Testing UDIF

Hipetesis 1 is a test UDIF. The results of the study are presented in Table 1 below.

Table 1. Uniform DIF Testing Results

Items	GLR Stat	p	UDIF	Items	GLR Stat	p	UDIF
1	0.328	0.955	No	21	1.719	0.633	No
2	0.491	0.921	No	22	1.410	0.703	No
3	1.090	0.780	No	23	3.590	0.309	No
4	0.239	0.971	No	24	0.192	0.979	No
5	15.676	0.001	Yes	25	4.902	0.179	No
6	8.567	0.036	Yes	26	0.507	0.917	No
7	0.956	0.812	No	27	12.114	0.007	Yes
8	1.457	0.692	No	28	1.587	0.662	No
9	1.883	0.597	No	29	0.726	0.867	No
10	1.633	0.652	No	30	0.768	0.857	No
11	0.616	0.893	No	31	3.268	0.352	No
12	4.426	0.219	No	32	1.742	0.628	No
13	1.236	0.744	No	33	7.762	0.051	No
14	1.611	0.657	No	34	2.375	0.498	No
15	3.404	0.334	No	35	7.485	0.058	No
16	42.569	0.000	Yes	36	3.094	0.377	No
17	0.535	0.911	No	37	5.345	0.148	No
18	7.815	0.050	Yes	38	11.073	0.011	Yes
19	2.267	0.519	No	39	5.865	0.118	No
20	4.309	0.230	No	40	1.037	0.792	No

Detection threshold: 7.8147 (significance level: 0.05)

Table I shows the results of testing UDIF of School Examination instrument in Soppeng. Based on these test results obtained the information that there are six items that are detected as Uniform DIF. The items are: 5 , 6 , 16 , 18 , 27 , and 38. These results can also be presented in the form of curves as follows:

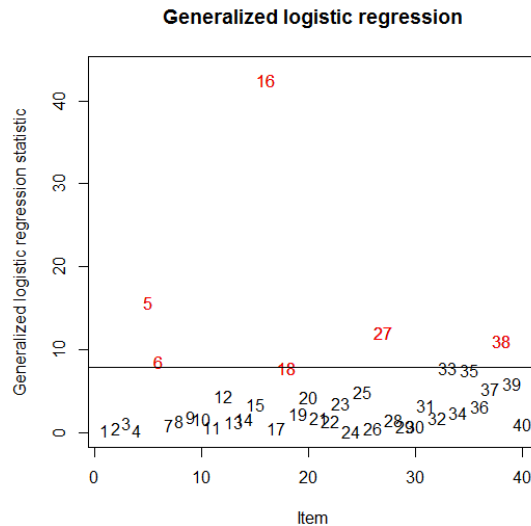
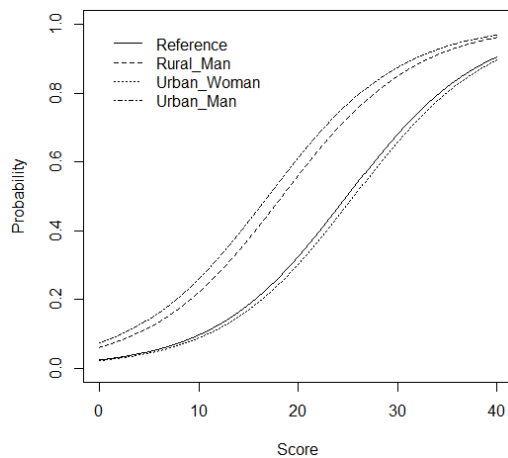
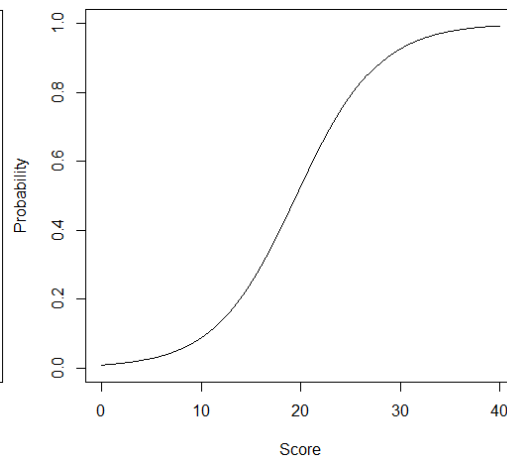


Figure 1. GLR curve Uniform DIF Testing Results

To further clarify the characteristics of the detected item contains Uniform DIF, then the following is presented an example of the characteristic curve for items detected UDIF and free UDIF based reference group and focal groups.

Figure 2. Item 5
(detected as uniform DIF)Figure 3. Item 8
(free of uniform DIF)

The plot shows that the item 5 is detected as UDIF items because of differences in the probability of success for several groups and the curve is uniform and does not intersect. Item 8 is an item that is free UDIF because the probability of success equal for all groups and the curves that do not intersect. In this case, other items that are detected as UDIF items are 6, 16, 18, 27, and 38.

B. Testing NUDIF

Hipetesis 2 is a test NUDIF. The results of the study are presented in Table 2 below.

Table 2. Nonuniform DIF Testing Results

Items	GLR Stat	<i>p</i>	NUDIF	Items	GLR Stat	<i>p</i>	NUDIF
1	1.705	0.636	No	21	3.262	0.353	No
2	6.501	0.090	No	22	0.386	0.943	No
3	4.827	0.185	No	23	4.354	0.226	No
4	0.962	0.811	No	24	1.796	0.616	No
5	1.413	0.703	No	25	0.969	0.809	No
6	0.686	0.877	No	26	1.893	0.595	No
7	2.076	0.557	No	27	1.866	0.601	No
8	5.914	0.116	No	28	0.455	0.929	No
9	3.135	0.371	No	29	3.444	0.328	No
10	3.305	0.347	No	30	1.953	0.582	No
11	0.761	0.859	No	31	0.636	0.888	No
12	0.948	0.814	No	32	0.266	0.966	No
13	4.035	0.258	No	33	1.984	0.576	No
14	0.489	0.921	No	34	0.670	0.880	No
15	1.153	0.764	No	35	2.590	0.459	No
16	34.114	0.000	Yes	36	1.974	0.578	No
17	5.163	0.160	No	37	0.036	0.998	No
18	0.775	0.856	No	38	0.515	0.916	No
19	8.886	0.031	Yes	39	4.182	0.243	No
20	1.087	0.780	No	40	2.257	0.521	No

Detection threshold: 7.8147 (significance level: 0.05)

Table 2 shows the results of testing NUDIF of School Examination instrument in Soppeng. Based on these test results obtained the information that there are two items that are detected as Nonuniform DIF. The items are: 16, and 19. These results can also be presented in the form of curves as follows:

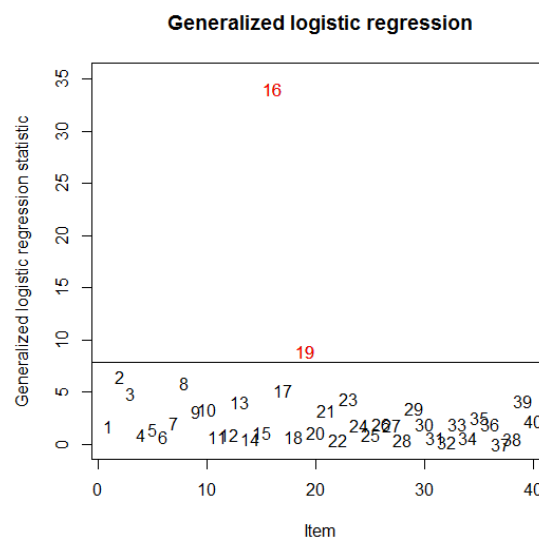


Figure 4. GLR curve Nonuniform DIF Testing Results

In addition, also presented the characteristics of the detected item contains Uniform DIF. The following are examples of item characteristic curve for item detected as NUDIF and item detected as free NUDIF based reference group and focal groups.

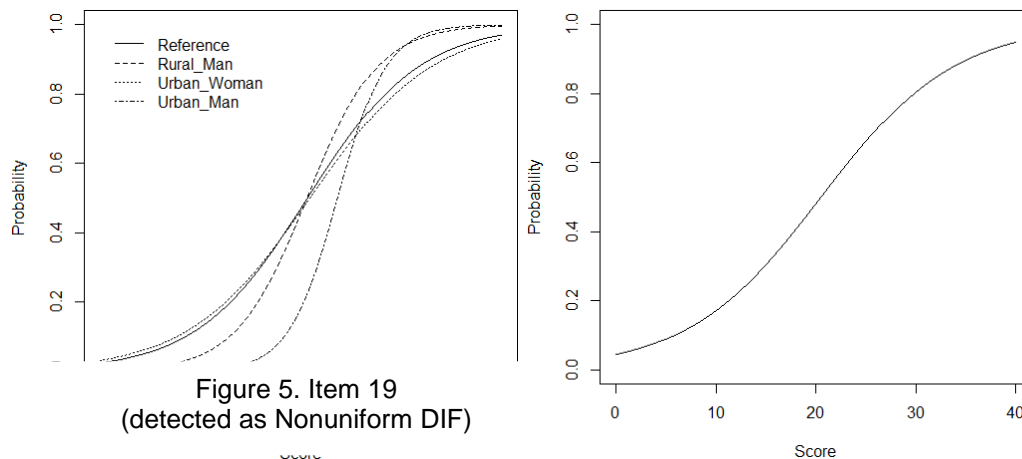


Figure 5. Item 19
(detected as Nonuniform DIF)

The plot shows that the item 19 is detected as NUDIF items because of differences in the probability of success for several groups and the curves intersecting. Item 21 is an item that is free NUDIF because the probability of success equal for all groups. In this case, other item that are detected as UDIF item is 16.

C. Testing Both (UDIF and NUDIF)

Hipetesis 3 is a test Both Type. The results of the study are presented in Table 3 below.

Table 3. DIF Both Type Testing Results

Items	GLR Stat	<i>p</i>	Both	Items	GLR Stat	<i>p</i>	Both
1	2.033	0.917	No	21	4.982	0.546	No
2	6.992	0.322	No	22	1.797	0.937	No
3	5.917	0.433	No	23	7.943	0.242	No
4	1.200	0.977	No	24	1.988	0.921	No
5	17.089	0.009	Yes	25	5.871	0.438	No
6	9.253	0.160	No	26	2.400	0.880	No
7	3.032	0.805	No	27	13.979	0.030	Yes
8	7.371	0.288	No	28	2.041	0.916	No
9	5.019	0.541	No	29	4.170	0.654	No
10	4.938	0.552	No	30	2.720	0.843	No
11	1.377	0.967	No	31	3.904	0.690	No
12	5.375	0.497	No	32	2.008	0.919	No
13	5.271	0.510	No	33	9.746	0.136	No
14	2.100	0.910	No	34	3.045	0.803	No
15	4.557	0.602	No	35	10.076	0.122	No
16	76.683	0.000	Yes	36	5.068	0.535	No
17	5.698	0.458	No	37	5.380	0.496	No
18	8.590	0.198	No	38	11.587	0.072	No
19	11.153	0.084	No	39	10.047	0.123	No
20	5.395	0.494	No	40	3.294	0.771	No

Detection threshold: 12.5916 (significance level: 0.05)

Table 3 shows the results of testing Both Type of School Examination instrument in Soppeng. Based on these test results obtained the information that there are three items that are detected as Nonuniform DIF. The items are: 5, 16, and 27. These results can also be presented in the form of curves as follows:

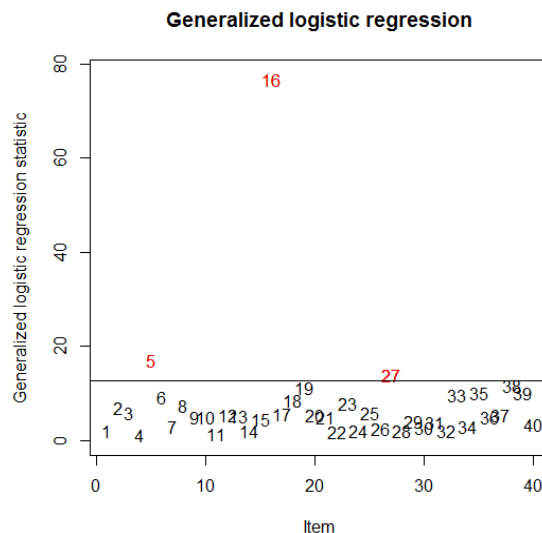


Figure 7. GLR curve Both Type DIF Testing Results

Like the previous test, to further clarify the characteristics of the detected item contains Both Type of DIF, then the following items are presented examples of characteristic curves for items that are detected as Both Type of DIF and free Both Type of DIF by reference group and focal groups.

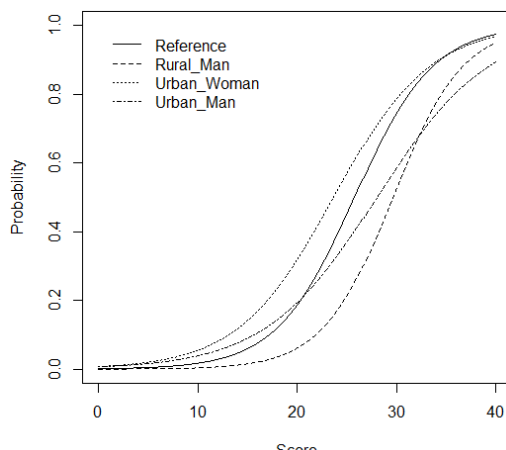


Figure 8. ITEM 27
(DETECTED AS BOTH TYPE OF DIF)

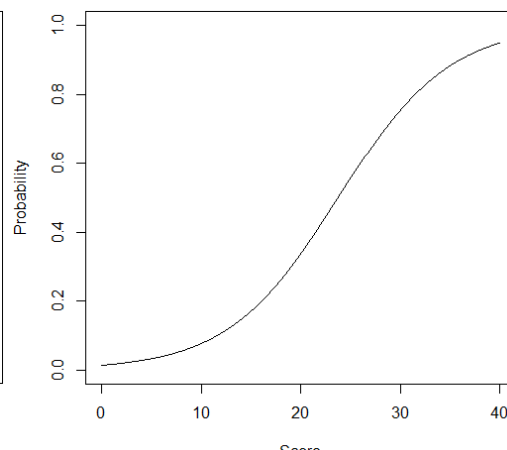


Figure 9. Item 30
(free of Both Type of DIF)

The plot shows that the item 27 is detected as Both Type of DIF items because the item contains UDIF and NUDIF. Item 30 is an item that is free DIF because the probability of success equal for all groups. In this case, other items that are detected as Both Type of DIF items are 5, and 16.

IV. CONCLUSIONS AND SUGGESTIONS

Based on the findings, the conclusions of this study are.

1. Instruments elementary school exams in Soppeng has six items containing Uniform DIF, namely: 5, 6, 16, 18, 27, and 38.
2. Instruments elementary school exams in Soppeng has two items containing Nonuniform DIF, namely: 16, and 19.

3. Instruments elementary school exams in Soppeng has three items containing Both Type of DIF, namely: 5, 16, and 27.

Based on these results, it is recommended the following.

1. To develop a quality instrument is recommended to perform the analysis item bias .
2. To perform DIF analysis, it is recommended to use the Generalized Logistic Regression Method.

REFERENCES

- [1] Jodoin, M. G., & Gierl, M. J. 2001. Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. *Applied Measurement in Education*, Vol. 14 No. 4, 329–349.
- [2] Magis, D., Raiche, G., Beland, S., & Gerard, P. A. 2011. Generalized Logistic Regression Procedure to Detect Differential Item Functioning Among Multiple Groups. *International Journal of Testing*, Vol. 11: 365–386.
- [3] Osterlind, S. J. 1983. *Test Item Bias. Series: Quantitative Applications in the Social Sciences*. California: Sage Publications.
- [4] Scheuneman, J. D., & Bleistein, C. A. 1999. *Item Bias: Advances in Measurement in Educational Research and Assessment*. New York: Pergamon.
- [5] Svetina, D., & Rutkowski, L. 2014. Detecting Differential Item Functioning using Generalized Logistic Regression in the Context of Large-scale Assessments. *Large-scale Assessment in Education Vol. 2 No.4*.
- [6] Zhang, Y. 2015. Multiple Ways to Detect Differential Item Functioning in SAS. *Educational Testing Service, Paper 2900*.
- [7] Zumbo, B. D. 1999. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF)*. Ottawa Canada: Directorate of Human Resources Research and Evaluation.

Effects of Complexity Matter and Grouping Students of the Statistics Analysis Capabilities

Ismanto

STAIN Kudus

ismanto_07@yahoo.co.id

Abstract- This study aimed to compare the ability of statistical analysis based on the complexity of the matter and grouping students of the study program. This type of research is ex-post facto of learning statistics that have been implemented in odd semester of academic year 2015/2016, as well as a quantitative approach. The data used in this study are the test results essay given by the two groups of students of the STAIN Kudus on the course for Islamic religious education (PAI) and primary teacher education (PGMI), i.e. the ability analysis of descriptive statistical, comparative and associative. Methods data collection also uses a semi-open questionnaire about learning statistics. Samples students determined by a Slovin formula that is equal to 155 of the 254 students of the two courses. Analysis of the data used is two-way ANOVA. The results of this study indicate that there are significant differences in statistical analysis capability is affected by the level of complexity in its statistical material with $F = 62.142$ and $\text{sig.} = 0.000$. There are significant differences statistical analysis capability is affected by the grouping of students based on courses with $F = 50.726$ and $\text{sig.} = 0.000$. As well as the higher complexity of the material will affect the ability of statistical analysis to students of both courses with $F = 1.914$ and $\text{sig.} = 0.149$. These results are supported by a group of students of primary teacher education said to be more independent in their learning compared to those students of Islamic religious education.

Keywords: *statistical analysis capabilities, the complexity of the matter, independence*

I. INTRODUCTION

A. The Background

Position statistics in the curriculum structure of Educational Department of STAIN Kudus as part of a clump of subjects of analysis, a quantitative analysis of basic subjects related to co-requisite courses, such as the development of systems of evaluation and quantitative research methods. Once the strategic position of these statistics requires the parties involved, a lecturer on the one side and students on the other, as subjects who are involved directly in the learning process both in the classroom and laboratory, to provide reinforcement for the roles of both parties. Lecturer facilitator function has an important role, from planning, process, and evaluation of learning statistics, so that not only the transfer of knowledge, but also as a unified whole value alignments in statistical learning itself.

B. Andragogy

Adult education is a form of learning experience needed by adult men and women in accordance with the interests and needs at the level of skills and knowledge are different and to support changes in the roles and responsibilities in life (Pannen and Malati, 1997: 7). Thus, the learning process in college, include one form of adult education, with the assumption that students are considered as adults. It means learning to students requires the student to make a creative, sensitive, aware and become members of the community play an active role in the development process.

There are several factors that affect student learning as an adult, which is a factor of freedom, responsibility, decision-making, self-direction, psychological, physical, memory, and motivation. Factor democratic freedoms, practical meaning and lead to solving the problem, while the responsibility of the student can be done through discussions, question and answer, independent tasks, and time deadlines. Decision-making in terms of linking the direction of the lessons learned to the needs of students, the experience and the good and bad ratings.

Psychological factors by humanistic, meaning that faculty and students can foster a sense of mutual need. Physical factors in a learning situation is comfortable, pleasant, relaxed, good room and good media placement. Factors memory by emphasize the inter-connection between knowledge and experience on cross time and novelty. Motivation factor for adult education by Houle (1961) were divided into three groups: goal-oriented with emphasis on the application and utilization of lessons, action-oriented with emphasis on learning and interaction among students, and oriented to the study of science itself for fun learning.

C. Active-Independent Learning

Self-learning is defined as an individuality business student autonomy to achieve academic competence (Kozma, Belle, & Williams, 1978). Independent study shows that students are not dependent on supervision (supervision) and faculty continuous guidance, but students also have the creativity and initiative of his own, and able to work independently with reference to the guidance obtained (Knowles, 1976). So the self-learning provides an opportunity for students in the development and improvement of capabilities and skills of students to conduct their own learning, does not depend entirely on teachers, classes, friends, and so on.

Active learning is an approach to learning management systems in ways that active learning toward self-learning and self-learning ability is the ultimate aim of active learning (Pannen and Sekarwinahyu, 1997: 3). In order to achieve this, learning activities are designed so that meaningful for students. Meaningful learning occurs when students actively participate in the learning process and finally be able to decide what to study and learn event. Active learning can be useful for development student learning ability in yourself and explore the potential of students and professors to jointly developing and sharing knowledge, skills and experience.

The development of active-independent learning skills can be done in two ways, namely lecturers provide students with cognitive strategies and guiding students through lectures contract. Armed with cognitive strategies and contract college students go through the process of self-directed learning. Students have the freedom to apply cognitive strategies own heuristic and requires creativity in the learning process to achieve the learning outcomes that have been agreed in the contract lectures.

D. Lecturer and Task

Tasks can make student learning more fun, effective, and efficient, because the task became to give students an opportunity to receive information in addition to the new, also apply, analyze and even evaluate such information. According Budiardjo (1997: 73-82), relating to the provision of the task, the lecturer has three roles: as a planner, facilitator, and evaluator. As a planner, lecturer is deciding the right task to do the student, so that the analysis of the assignment has no basis in initial skills, curiosity, and the time it takes students to complete. As a facilitator, lecturer provided direction active students to learn independently, i.e. learn to describe, analyze, compare, make an analogy, infer, synthesize and / or evaluating a concept or principle. Then as facilitators, lecturers can provide the type of assistance such as task remedial work to correct mistakes students, the feedback in the form of comments and suggestions to the member a chance of finding answers, and provide reinforcement in the form of praise or appreciation of the development tasks that have been completed.

E. Cognitive Domain of Statistics

Cognitive domains accord to Bloom's taxonomy has six levels of thinking, namely knowledge, comprehension, application, analysis, synthesis, and evaluation). This paper specifically, parse domain of cognition that can be developed in the context of statistical learning in STAIN Kudus, include knowledge, comprehension, application and analysis. Four levels of proficiency level bind one another to explain the hierarchy of learning statistics that include descriptive statistical analysis, comparative and associative. According Retnawati (2015: 7-9), said fourth operational work domain of cognition used in this study to unravel the hierarchy of statistical learning as a verb describes, counting, registering, tabulate, drawing, distinguish, associate, calculating, simulating, predicting, analyzing, test, and concluded.

Descriptive statistical analysis, generally receipts verb describes, counting, registering, tabulate, and drawing. Meanwhile, comparative analysis and associative uses the verbal differentiate, associate, calculate, simulate, predict, analyze, test, and concluded. Comparative analysis although his position is almost equivalent to the associative analysis, but both in the material structure has different levels, where the material one-way ANOVA as a matter of

hypothesis testing comparative hypothesis, and will be needed again when the associative hypothesis testing.

F. Grouping students

Grouping students in this study refers to the opinion of Thompson (1967: 70) states that all organizations face the task of heterogeneous environments tried to identify homogeneous segments and establish structural units to deal with each issue. Statement Thompson gives the sense that the organization can achieve more efficient when allocating separate tasks for special subunits, and this research see that a college in it have working units with the intent to carry out optimally every job he had. Educational Department of STAIN Kudus as one of the public services unit specifically in the field of education to every student, and it has several sub-units called the study program.

G. Research problem

Based on the above background, this study to formulate the problem of is there any difference in the ability of statistical analysis based on the complexity of the matter and grouping students of the Educational Department of STAIN Kudus.

H. Objectives

This study aimed to compare the ability of statistical analysis to consider the structure of matter descriptive statistical analysis, comparative and associative, and grouping students of the study program in PAI and PGMI.

I. Significance of research

The benefits of this research to prove that the hierarchical structure of statistical material in the learning process and the need for statistics in the structure of the curriculum for a course PAI and otherness in primary impact analysis capabilities in every part.

II. RESEARCH METHOD

This study uses a quantitative approach, meaning that the data involved in testing the hypothesis in this study of the data statistical analysis capabilities, the student response data from the statistical material about the description. The population involved in this study were students of the study program in PAI and PGMI a number of 254 students, while the sampling is done randomly by simple random sampling technique, as well as the size of the sample using the formula Slovin number of 155 students were involved in the test and 243 students involved charging non-test. The instrument of this research is the statistical test contains descriptive statistical analysis, comparative and associative, and supported by a semi-open questionnaire instrument containing grains surrounding the teaching and learning of statistics, where respondents can specify more than one option available answers as appropriate. Source of the data of this study is the result of statistical learning student of PAI and in primary odd semester of the 2015/2016 academic year. Technique of the data analysis used in this study is a two-way ANOVA (Budiyo, 2009: 228) aided by SPSS version 16.0.

III. RESEARCH FINDINGS AND DISCUSSIONS

The result of this research is descriptive and comparative analysis results. Descriptive analysis, statistical analysis capabilities include fly behind in terms of courses and complexity of the matter, as follows:

Table 1. Descriptive Statistics of the complexity of the matter

Complexity Matter	Grouping students	Mean	Std. Deviation	N
descriptive analysis	PAI	47.828	23.4097	61
	PGMI	61.543	25.1263	94
	Total	56.145	25.2973	155
comparative analysis	PAI	64.377	33.8497	61
	PGMI	88.181	17.8265	94
	Total	78.813	27.8287	155
associative analysis	PAI	72.705	22.9596	61
	PGMI	88.245	19.3678	94
	Total	82.129	22.1341	155

From Table 1 above it appears that the scale of 100 votes, the ability of descriptive statistical analysis, comparative and associative respectively 56.145, 78.813 and 82.129. Descriptive analysis capabilities has the lowest average period of adjustment due to the statistical study of general teaching subjects other than statistics, so that the average student still find it difficult to apply statistical concepts in the field of education, especially related to quantitative data. While the average for all three statistical capability has increased, meaning that after a few times to follow and observe statistical learning materials, student of Educational Department can receive well, both in substance and interrelation between learning material each meeting.

Table 2. Descriptive Statistics of the study program

Complexity Matter	Grouping students	Mean	Std. Deviation	N
Total	PAI	61.637	28.9771	183
	PGMI	79.323	24.4315	282
	Total	72.362	27.6704	465

While based on Table 2, it appears that the statistical analysis capabilities with three items, namely descriptive statistical analysis, comparative and associative had a mean 61.637 for the course PAI and 79.323 for the study program in primary. This means that the ability of statistical analysis in primary students of higher than PAI. This is supported also by the activeness and independence of the study both groups (in the form of liveliness lectures, learning styles, application data, and assignment), it appears that the students of PGMI are more active compared with PAI-independent, with the proportion of each response non-test 70.23 % and 56.86%.

Meanwhile, comparative analysis, including differences in statistical analysis capabilities that are affected by the fly behind the program of study and/or complexity of the matter, as follows:

Table 3. Summary Table of Two-Way ANOVA

Dependent Variable: Statistics Analysis Capabilities

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
complex	56673.997	2	28336.999	50.726	.000
group	34714.472	1	34714.472	62.142	.000
complex * group	2138.419	2	1069.209	1.914	.149
Total	2790148.250	465			

a. R Squared = .278 (Adjusted R Squared = .270)

Estimated Marginal Means of Statistics Analysis Capabilities

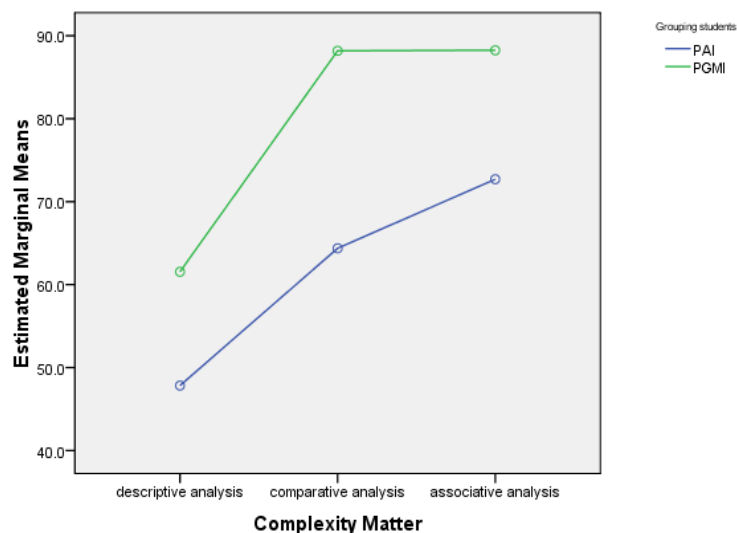


Figure 1. Profile of interaction between the complexity of the matter and grouping students

Table 3 above explains that there are significant differences in terms of statistical analysis capabilities of the complexity of the material descriptive statistical analysis, comparative and associative, it can be seen the value $F = 50.726$ with $\text{sig.} = 0.000$. Can be explained also that there are significant differences as well as statistical analysis capabilities in terms of grouping PAI and PGMI courses to the value of $F = 62.142$ and $\text{sig.} = 0.000$. As well as the higher complexity of the material will affect the ability of statistical analysis to students of both courses with $F = 1.914$ and $\text{sig.} = 0.149$, this can be seen profile interaction below, which states that there is no interaction between the complexity of the matter of statistics by grouping students in statistical analysis capabilities.

It turned out that over time the statistical learning process, student of educational Department of the STAIN Kudus both PAI and PGMI courses has significantly improved its statistical analysis capabilities, graph above shows an increase for both courses.

IV. CONCLUSIONS AND SUGGESTIONS

Based on the analysis above study, it can be concluded that the higher complexity of the material will affect the ability of statistical analysis student of PAI and PGMI. Proof of the above is also supported that group of students of PGMI is said to be more active in the self-study compared groups of students of PAI, it can be seen in profile the interaction of both courses. While the suggestion of this research is the need to learn a longer time than originally only one semester, to be more mature in building and developing the scientific foundations in terms of quantitative analysis capabilities.

REFERENCES

- [1] Budiyo. (2009). *Statistika untuk penelitian*. Surakarta: UNS Pess.
- [2] Heri Retnawati. (2015). *Validitas, reliabilitas & karakteristik butir (panduan untuk peneliti, mahasiswa, dan psikometrian)*. Yogyakarta: Parama Publishing.
- [3] Houle, C. O. (1961). *The inquiring mind*. Madison: University of Wisconsin Press.
- [4] Knowles, M. S. (1975). *Self-directed learning: A guided for learners and teachers*. New York: Cambridge Adult Education.
- [5] Kozma, R. B., Belle, L. W., & Williams, G. W. (1978). *Instructional techniques in higher education*. Educational Technology. Englewood Cliffs, New Jersey: Educational Technology Publications.
- [6] Lily Budiardjo. (1997). Pendidikan orang dewasa. Dalam PAU-PPAI-UT. *Mengajar di perguruan tinggi: Program applied research*. Jakarta: Direktorat Jenderal Pendidikan Tinggi, Departemen Pendidikan dan Kebudayaan.
- [7] Paulina Pannen & Ida Malati S. (1997). Pendidikan orang dewasa. Dalam PAU-PPAI-UT. *Mengajar di perguruan tinggi: Program applied research*. Jakarta: Direktorat Jenderal Pendidikan Tinggi, Departemen Pendidikan dan Kebudayaan.
- [8] Paulina Pannen & Mestika Sekarwinahyu (1997). *Pendidikan orang dewasa*. Dalam PAU-PPAI-UT. *Mengajar di perguruan tinggi: program applied research*. Jakarta: Direktorat Jenderal Pendidikan Tinggi, Departemen Pendidikan dan Kebudayaan.
- [9] Thompson, J. D. (1967). *Organizations in action*. New York: McGraw-Hill

Construct Validity of the TGMD-2 in 7–10-Year-Old Surakarta Children with Mild Mental Disorder

Ismaryati

Department of Sport and Health Education, Faculty of Teacher Training and Education
Sebelas Maret University, Jl. Ir. Sutami 36 A Kentingan, Surakarta Indonesia 57126
e-mail: iis_050563@yahoo.co.id

Abstract- The purpose of this study was to assess the Construct Validity of the TGMD-2 on Surakarta children with Mild Mental Disorder. The total sample consisted of 99 children aged 7-10 years of which 67 were boys and 32 were girls, represents 80% of all students in Surakarta extra ordinary elementary school. The two-factor structure of the Test of Gross Motor Development-2 was tested using confirmatory factor analysis with maximum likelihood estimation to compute parameter estimates and to select the appropriate item for each factor. The overall fit of the data to the model was initially based on the chi-square statistic. Preliminary results suggested a poor fit $\chi^2 = 83.772$, $DF = 53$, $p = 0.004$. Nevertheless, the chi-square statistic is somewhat sensitive to sample size (Byrne, 1989), so that multiple fit indices were also applied. For this study as well as results from the TGMD-2. A value for the relative chi-square close to 1 indicates that the model is correct. In the present study, the relative chi-square was 1.58. Values for GFI and AGFI should be .90 or above .80, respectively. For this study, GFI is equal to .88 and AGFI is equal to .82. The model thus appears to fits sufficiently.

Key words: Confirmatory factor analysis; test of gross motor development, Mild Mental Disorder

I. INTRODUCTION

Fundamental movement skills (FMS) refer to goal-directed movement patterns (Burton & Miller, 1998). FMS consist of locomotor skills (walking, running, hopping, galloping, jumping, sliding, and leaping) and object control skills (throwing, catching, striking, pulling, and pushing, bouncing, kicking) performed in the bipedal position (Burton & Miller, 1998). Fundamental movement skills allow children to move through space (Zittel, 1994) and permit them how to respond in a suitable way to a variety of stimuli (Krebs, 2000). The proficiency of gross motor skills is a prerequisite for children to experience success and enjoyment in organized and unorganized movement activities (Woodard & Surburg, 2001). Mastery of these skills is a prerequisite to the successful introduction of specific sports and games (Burton & Miller, 1998;) with practice and instruction being crucial to their development (Gallahue & Ozmun, 1998). Inadequate motor skill proficiency at an early age can negatively influence competence in physical and motor activities in later life (Jurimae & Jurimae, 2000).

Motor assessment is particularly important during the preschool and school years as it monitors developmental changes and identifies developmental delays (Gallahue & Ozmun, 1998). Therefore, information from accurate gross motor skills assessment can be profitably used by physical educators and sports administrators. They can use this information to examine children's gross motor skills performance, provide information for educational programs and instructions, assess the gross motor development status of individuals, and assess the effectiveness of motor development programs (Burton & Miller, 1998). Numerous instruments are available to assess fundamental motor skills, within existing physical education time (Zittel, 1994).

The current TGMD-2 is a process-oriented assessment accompanied by criterion-referenced and norm-referenced interpretations. It was designed to evaluate the gross motor skill process or pattern exhibited by children with chronological age equivalents of 3 to 10 years of age. In developing the TGMD-2, the following changes and improvements were made: (a) the norms provided, cover chronological age per six months in preference of twelve months; (b) reliability was improved; (c) separate norm tables for the object control subset for each gender became available; (d) validity was tested in depth; (e) the item skipping was eliminated from the locomotor

subset, and the item of underhand roll was added to the object control subset; (f) the criteria of some specific items were adjusted; (g) tests were now performed only twice and not three times; (h) norm tables are now available to convert raw scores (a) to percentiles according to age and performance and (b) in age equivalents.

Construct validity is the degree to which a test measures what it claims, or purports, to be measuring (Cronbach and Meehl: 1955). Construct validity is the appropriateness of inferences made on the basis of observations or measurements (often test scores), specifically whether a test measures the intended construct. Construct-identification validity refers to the degree to which the underlying traits of a test can be identified and the extent to which these traits reflect the model on which the test is based (Anastasi & Urbina, 1997).

Evidence of the validity of the TGMD-2 was reported in its test manual (Ulrich, 2000). The TGMD-2 was validated on 1,208 American children. The validity of this test instrument was examined via exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). In terms of EFA, Ulrich (2000) utilized principal component analysis with promax rotation to examine the underlying structure of the TGMD-2. The EFA results identified two factors, with Strike, Dribble, Catch, Kick, Throw, and Roll loaded on the first factor and Hop, Slide, Gallop, Jump, Leap, and Run loaded on the second factor. Hop had the highest factor loading in the Locomotors subtest (.70), while Leap had the lowest (.49). For the Object Control subtest, Strike and Roll demonstrated the highest factor loading (.75), while the lowest was Catch (.57). CFA was utilized by Ulrich (2000) to further explain the underlying structure of the TGMD-2. The fit indexes (goodness-of-fit index [GFI] and adjusted GFI [AGFI]) indicated that the two factor model of the TGMD-2 produced a good approximation to data (GFI=.96, AGFI=.95).

Motor function is developmental in nature; performance on the TGMD-2 should be strongly correlated with chronological age (Ulrich, 2000). In children at risk for developmental disorders, competence in fundamental motor skills is limited, and development may be less predictable (Goodway & Branta, 2003; Goodway, Crowe, & Ward, 2003;). These children exhibit a delay both in gross and fine motor skills (DiRocco, Clark, & Phillips, 1987; Goodway & Branta, 2003; Goodway, Crowe, & Ward, 2003; Krebs, 2000) and consistently score lower than children without mental disorder on measurements of strength, endurance, agility, balance, running, speed, flexibility, coordination, and reaction time (Gallahue & Ozmun, 1998; Krebs, 2000; Shephard, 1990); however, with additional practice, performance can be improved (Zittel & McCubbin, 1996).

TGMD-2 has only been applied to children without intellectual disability. To our knowledge no studies have yet been conducted to examine the differences in performance between children without mental disorder and children with mild mental disorder in fundamental movement skills using the TGMD-2. To our knowledge no studies have yet been conducted to examine the differences in performance between children without intellectual disability and children with mild mental disorder in fundamental movement skills using the TGMD-2 in Surakarta.

II. RESEARCH MEHOD

Ninety-nine Mild Mental Disorder children aged 7-10 years participated in the study. Children were included in this study if they (a) attending Extraordinary elementary schools state and private in Great Solo and (b) had no physical disability hindering participation in the test. Schools for special education in Indonesia are organized into five types. Type A for children with visually impaired, Type B for children with hearing impaired, Type C for children with mentally disabled, Type D for children with physical impairment, Type D for children with emotional behavior disorder, and special schools with autism. Schools for special education in Indonesia, especially type C grouped into two type; educable (C) and Trainable (C1). Type C is children with mild mental disorder without physical disability hindering participation in the test.

Mean age of participants was 8 years, 10 months ($SD = 1$ year, 9 months), ranging from 7 years to 10 years, 9 months old. Mean age of 67 boys was 9 years, 8 months ($SD = 1$ year, 2 months) and 8 years, 8 months ($SD = 10$ months) for the 32 girls. Seven children were excluded from the original sample due to a total intelligence quotient that was too low or too high, according to the definition of mild intellectual disability of the American Association on Intellectual and Developmental Disabilities (AAIDD, 2007). Mental retardation is a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills (AAIDD, 2007). Participants' total Intelligence Quotient (TIQ) scores ranged from 52 to 70 and were measured by 8 different Intelligence tests

used by the various child guidance services; therefore, the mean and standard deviation for the total-IQ scores were not calculated. The scales used were the Wechsler Intelligence Scale for children (Wechsler, 1992).

The TGMD-2 evaluates fundamental movement skills of typically developing children between 3 and 10 years of age (Ulrich, 2000). The test is divided into two subsets: the locomotor skill subset and the object control skill subset. The locomotor skill subset consists of subtests for run, gallop, jump, slide, hop, and leap. There are four performance criteria for the first four skills, while hopping and leaping are judged on five and three criteria, respectively. The object control subset is also composed of six skills: striking a stationary ball, stationary dribble, catch, kick, overhand throw, and underhand roll. Stationary dribble, kick, overhand throw, and underhand roll are judged on four criteria. Striking a stationary ball has five and catching has three performance criteria.

Each skill was performed twice, and a single rater gives a score of 1 or 0 for each criterion. The examiner then totals the scores for each criterion for the two trials of each skill to obtain a raw skill score. All skill scores of each subset (locomotor and object control) are then added to form a raw subset score. Each raw subset score for locomotor skill as well as object control skill are converted to a standard score ($M = 10$, $SD = 3$) using norm tables (Ulrich, 2000). Finally both subset standard scores are combined and converted according to norm tables to an overall Gross Motor Quotient (GMQ; $M = 100$, $SD = 15$; Ulrich, 2000).

Statistical analyses were performed using Statistica 6.00 (Statsoft, 2004). Normality was checked using the Kolmogorov-Smirnov test. To investigate the construct validity of the TGMD-2, confirmatory factor analysis was used. Maximum likelihood on confirmatory factor analyses using five fitting indexes was performed to test the goodness of fit of the TGMD-2 skills to the locomotor and object control subsets. For comparison of results, the same four indexes used by Ulrich (2000) were taken.

Analysis of variance was used to examine changes with age in locomotor and object control subset raw scores. Effect size (η^2) is reported when significant results were obtained. To determine the differences in performance between U.S. children without intellectual disability from the normative sample (Ulrich, 2000) and Surakarta children with mild mental disorder, used a Z-test as the standard deviation for the population is known (Clark-Carter, 2004).

Pearson product moment correlations were calculated to examine the relationship between each item within the two subsets as well as the subset scores with the total score. The correlation between the standard scores in both subsets and the Gross Motor Quotient was also determined. Three aspects of reliability of the TGMD-2 were checked: (a) internal consistency reliability using Cronbach's coefficient Alphas, and (b) inter-rater reliability was determined using Pearson product moment correlations (Ulrich, 2000). Scorer error can be reduced with the use of clear administration procedures, detailed guidelines governing scoring, and practice (Ulrich, 2000). In this case, the protocols were selected randomly and completed independently by two investigators. The raw scores were converted to standard scores and then correlated. Test-retest values were calculated by means of the Spearman correlation.

III. RESEARCH FINDING AND DISCUSSION

The Kolmogorov-Smirnov test for normality showed no significant p -values for the locomotor tests ($\max D = 0.112$, $p < .15$), object control tests ($\max D = 0.109$, $p < .20$), and the overall Gross Motor Quotient ($\max D = 0.117$, $p < .15$). The p -values for both subtests and the overall GMQ were not significant. The test results were considered to be normally distributed and therefore parametric statistics were used.

Initially, Ulrich's (2000) two-factor model was examined in this study. It is hypothesized that one factor "locomotion" is comprised of six items (run, gallop, leap, hop, jump, and slide) while the remaining variables (striking a stationary ball, stationary dribble, catch, kick, overhand throw, and underhand roll) were related to the factor "object control skills." To test this hypothesis, a confirmatory factor analysis was used. The overall fit of the data to the model was initially based on the chi-square statistic. Preliminary results suggested a poor fit ($\chi^2 = 83.772$, $DF = 53$, $p = 0.004$). Nevertheless, the chi-square statistic is somewhat sensitive to sample size (Byrne, 1989), so that multiple fit indices were also applied. Table 1 presents the results for three indices of fit: the relative chi-square (Wheaton, Muthén, Alwin, & Summers, 1977), the Goodness of fit index (GFI; Jöreskog & Sörbom, 1993), and the adjusted goodness of fit index (AGFI; Jöreskog

& Sörbom, 1993) for this study as well as results from the TGMD-2 manual (Ulrich, 2000). A value for the relative chi-square close to 1 indicates that the model is correct (Byrne, Shavelson, & Munthén, 1989). In the present study, the relative chi-square was 1.58. Values for GFI and AGFI should be .90 or above .80, respectively. For this study, GFI is equal to .88 and AGFI is equal to .82. The model thus appears to fit sufficiently.

Table 1. Goodness Of Fit Indexes Of The Tgmd-2

Chi-square	Current study	Ulrich (2000)
	83.77	280.3
Df	53	53
χ^2/df	1.58	5.29
GFI	.88	.96
AGFI	.82	.95

When comparing the standard scores of our participants to the standard scores of the normative United States group (Ulrich, 2000; see Table 2), the results of children with mild mental disorder were significantly lower. Large effect sizes were also found ranging from $ES = 1.22$ to $ES = 1.57$, indicating a large difference. All items of the locomotor subtests were significantly but moderately correlated with the locomotor subset score ($p < .05$; range $r = .48$ to $r = .67$). In addition, all six items included in object control subset were moderately correlated with object control subset raw score ($p < .05$; range $r = .36$ to $r = .76$). A significant correlation between object control and locomotor standard scores was also found ($r = .54$; $p < .05$). In addition, Gross Motor Quotient was highly correlated with both the object control subset standard score ($r = .89$; $p < .05$) and the locomotor subset standard score ($r = .86$; $p < .05$).

Table 2. Standard Scores And Gross Motor Quotient Scores In The United States ($N = 1208$) And Surakarta ($N = 99$) Group Of Children

	U.S group		Surakarta group				
	M	SD	M	SD	Z-value	p-value	ES
Locomotor	10	3	6.38	2.41	-12.007	< 0.001	1.22
Object control	10	3	5.86	2.69	-13.731	< 0.001	1.39
GMQ	100	15	76.67	13.46	-59.936	< 0.001	1.57

In the TGMD-2 manual (Ulrich, 2000), three types of error variance are described: internal consistency reliability, test-retest reliability, and inter-rater reliability. For Ulrich's data regarding internal consistency, Cronbach's Alpha coefficient was higher than $\alpha = .85$. object control $\alpha = .88$, and GMQ $\alpha = .91$. Correlation coefficient for the test-retest was .88 for locomotor, .93 for object control to and for the total test .96. Inter-rater reliability was .98 for all three aspects. In this study, Cronbach Alpha for the locomotor subset was $\alpha = .82$, for the object control subset $\alpha = .86$ and for the gross motor quotient $\alpha = .90$. Furthermore, inter-rater reliability was extremely high, with agreement between the first and the second scorer in the locomotor subset ($r = 1.00$; $p < .05$), the object control subset ($r = 1.00$; $p < .05$) and for the GMQ ($r = 1.00$; $p < .05$) scores. The test-retest correlation for the locomotor subset was .90, for the object control subset .92 and for the total test .98, see table 3.

Table 3. Reliability Coefficients Of The Tgmd-2

Test item	Current study	Ulrich (2000)
Internal consistency		
Locomotor scale	.82	.85
Object control scale	.86	.88
Gross motor quotient	.90	.93
Test-retest reliability		
Locomotor scale	.90	.88
Object control scale	.92	.93
Gross motor quotient	.98	.96
Inter-rater reliability		
Locomotor scale	1.00	.98
Object control scale	1.00	.98
Gross motor quotient	1.00	.98

IV. DISCUSSION

The purpose of this study was to assess the Construct Validity of the TGMD-2 on Surakarta children with Mild Mental Disorder. Validity was measured using confirmatory factor analysis. The confirmatory factor analysis, which was performed to test the goodness-of-fit of the items to the locomotor subset and object control subset, detected similar values for chi-square and relative chi-square as those given by Ulrich (2000). In this study, GFI and AGFI approached desirable levels of .90. The model thus appears to fit sufficiently. On the other hand, the group of participants here was rather small, which could have an impact on the results (Byrne, 1989). Further research utilizing a greater number of participants is needed.

The correlation for his population from the United States varied from $r = .69$ ($p < .05$) to $r = .72$ ($p < .05$) for the locomotor subset and $r = .71$ ($p < .05$) to $r = .75$ ($p < .05$) for the object control subset. Results this study showed a significant correlation between the age and object control subset scores ($r = .25$; $p < .05$). Ulrich (2000) indicated that both subsets were strongly related to the chronological age in children with typical developing children. This may be caused a disproportion between the ages in our group and a rather small sample. The total group consists out of 67 boys and 32 girls. On the other hand, in the age band of seven years there are 16 boys versus 5 girls and in the age band of 8 years there were 10 boys versus 13 girls. Eggert (1971) and Bruininks (1978) have concluded previously that the correlation between age and motor abilities was lower with children with mild mental disorder than in typically developing children.

As expected, children without disabilities from the Ulrich sample (2000) performed better than the Surakarta children with mild mental disorder. A large effect size (1.22–1.57) was found in both standard subset raw scores and Gross Motor Quotient. Norms for children with Down syndrome are also available in the TGMD-2 manual (Ulrich, 2000). These norms were also much lower than those of children without disabilities. Children with intellectual disability have previously been reported to exhibit a reduced rate of gross and fine motor skill development, compared to typical children of the same chronological age (Bruininks & Chvat, 1990; Newell, 1997; Vermeer & Davis, 1995; Wade 1986).

It was expected that there would be a significant correlation between the raw scores of the same subset skill score. Results of this study showed that all items were significantly correlated ($p < .05$) with their raw total scores and with the subset skill score. Only leap ($r = .49$, $p < .05$) and slide ($r = .48$, $p < .05$) for the locomotor subset were moderately correlated with the total subsets score, with hopping having the highest correlation ($r = .67$, $p < 0.01$). For the object control subset, the highest correlation was showed from dribbling ($r = .74$, $p < .05$) and the lowest for catching ($r = .36$, $p < .05$). This is in agreement with Ulrich (2000) who reported that tests must be composed of items that have high inter-correlation to show construct validity.

The object control subset as well as the locomotor subset was also significantly correlated with Gross Motor Quotient: $r = .89$ ($p < .05$) and $r = .86$ ($p < .05$), respectively. Similar results were also reported by Ulrich (2000). In this study the correlation of the Gross Motor Quotient with both subsets was significant, respectively ranging from $r = .92$ ($p < .05$) to $r = .88$ ($p < .05$). Additional evidence for the validity of the test was provided by the moderate correlations between the results on both subsets and the total score. In this study, both subsets were significantly correlated with one another, but to a moderate degree ($r = .54$; $p < .05$). This was expected, as the two subsets measured different aspects of fundamental movement skills. Furthermore, when this correlation is high, the two subsets would be measuring the same thing. If, on the other hand, the correlation is low, subsets would be considered to be measuring irrelevant abilities.

Reliability was examined using three methods: (a) internal consistency, (b) inter-rater reliability, and (c) intra-rater reliability. Similar values were reported by Ulrich (2000), in this study Cronbach's alpha coefficient for both subsets exceeded $\alpha = .82$ and it even reached $\alpha = .90$ for Gross Motor Quotient. This indicates that the TGMD-2 is a reliable tool for measuring fundamental movement skills in children with mild mental disorder

This result is clear administration procedures because examination of inter-rater reliability coefficients are usually made after video analysis or performed twice, we investigated this in relation to the TGMD-2 manual (Ulrich, 2000) by omitting the initial calculations (first observer) concerning the raw scores and rescoring the test for a second time (second observer). That was the reason inter-rater reliability reached the absolute value of $r = 1.00$ for all subsets as well as for the Gross motor Quotient. Finally, we found high test-retest values for an randomly selected

group of children ranging from $rs = .90$ for the locomotor subset, to over $rs = .92$ for the object control subset and $rs = .98$ for the Gross Motor Quotient

V. CONCLUSIONS AND SUGGESTIONS

The findings of this study provide support for the two-factor structure of the TGMD-2 originally developed by Ulrich (2000). Additionally, comprehensive information concerning the validity and reliability of the TGMD-2 is provided. The findings provide further evidence of the usability of the TGMD-2 when attempting to assess the gross motor skill performance of children with mild mental disorder. To increase the value of the object control subtest correlations suggested to use the more number of test participants.

REFERENCES

- [1] American Association on Intellectual and Developmental Disabilities (AAIDD). (2007). *Definition of Mental retardation*. www.aamr.org, updated July 5, 2007.
- [2] Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th^{ed.}). Upper Saddle River, NJ: Prentice-Hall.
- [3] Bruininks, R.H. (1978). *Bruininks-Oseretsky Test of Motor Proficiency. Examiners Manual*. Circle Pines, MN: American Guidance Service, Inc.
- [4] Bruininks, R.H., & Chvat, M. (1990). Research on the motor proficiency of persons with mental retardation. In H. Van Coppenolle & J. Simons (Eds.), *Better by Moving. Proceedings of the 2nd International Symposium Psychomotor Therapy and Adapted Physical Activity*. (pp. 43-69). Leuven, Belgium: Acco.
- [5] Burton, W.A., & Miller, E.D. (1998). *Movement skill assessment*. Champaign, IL: Human Kinetics.
- [6] Byrne, B.M. (1989). *A primer of LISREL: Basic application and programming for confirmatory factor analytic models*. New York: Springer-Verlag.
- [7] Byrne, B.W., Shavelson, R.J., & Munthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- [8] Clark-Carter, D. (2004). *Quantitative psychological research. A student's handbook*. New York: Psychological Press.
- [9] Cronbach, L. J., Meehl, P.E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin* 52 (4): 281-301
- [10] DiRocco, J.P., Clark, E.J., & Phillips, J.S. (1987). Jumping coordination patterns of mildly mentally retarded children. *Adapted Physical Activity Quarterly*, 4, 178-191.
- [11] Eggert, D. (1971). *Lincoln-Oseretsky-Skala. Manual*. Weinheim, Germany: Beltz Test Gesellschaft.
- [12] Eichstaedt, B.C., & Lavay, W.B. (1992). *Physical activity for individuals with mental retardation*. Champaign, IL: Human Kinetics.
- [13] Evagelinou, C., Tsigilis, N., & Papa, A. (2002). Construct validity of the test of Gross Motor Development: A cross-validation approach. *Adapted Physical Activity Quarterly*, 19, 483-495 Flemish Ministry of Education (2005-2006). Available at: http://www.ond.vlaanderen.be/onderwijsstatistiek/2005-2006/jb05-06/jb05-06-1_hdst3/.indd.pdf.
- [14] Gallahue, L.D., & Ozmun, C.J. (1998). *Understanding motor development. Infants, children, adolescents, adults*. Boston: McGraw-Hill.
- [15] Goodway, D.J., Crowe, H., & Ward, P. (2003). Effects of motor skill instruction of fundamental motor skill development. *Adapted Physical Activity Quarterly*, 20, 298-314.
- [16] Goodway, J.D., & Branta, C.F. (2003). Influence of a motor skill intervention on fundamental motor skill development in disadvantaged preschool children. *Research Quarterly for Exercise and Sports*, 74, 36-46.
- [17] Haubenstricker, J.L., & Seefeldt, V. (1986). Acquisition of motor skills during childhood. In V. Seefeldt (Ed.), *Physical activity and well-being* (pp 42-102). Reston VA: AAHPERD.
- [18] Holland, B.V. (1987). Fundamental motor skill performance of non-handicapped and educable mentally impaired students. *Education and training of mentally retarded*, 22, 197-204.
- [19] Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structure Equation Modeling With the SIMPLIS Command Language*. Chicago: Scientific Software International
- [20] Jurimae, T., & Jurimae, J. (2000). *Growth, physical activity, and motor development in prepubertal children*. Boca Raton, FL: CRC Press LLC
- [21] Karabourniotis, D., Evagelinou, C., Tzetzis, G., & Kourtessis, T. (2002). Curriculum enrichment with self-testing activities in development of fundamental movement skills of first-grade children in Greece. *Perceptual Motor Skills*, 94, 1259-1270.
- [22] Kim, J. (2003). *The effects of a physical education program on the standing long jump performance of preschool-aged children with cognitive delays*. Eugene, OR: Kinesiology Publications.
- [23] Krebs, P. (2000). Mental retardation. In J.P. Winnick (Ed.), *Adapted Physical Education and Sport*. (pp 111-126). Champaign, IL: Human Kinetics.
- [24] Langendorfer, S. (1986). Review of the test of Gross Motor Development. *Adapted Physical Activity Quarterly*, 3, 186-190.
- [25] Malina, R.M., & Bouchard, C. (1991). *Growth, Maturation and Physical Activity*. Champaign, IL: Human Kinetics.
- [26] Nelson, J.K., Thomas, J.R., Nelson, K.R., & Abraham, P.C. (1986). Gender. In: H. De Fever,
- [27] Newell, K.M. (1997). Motor skills and mental retardation. In W.E. Maclean (Ed), *Ellis' handbook of mental deficiency, psychological theory and research*. (pp. 275-308).

- [28] Shephard, J.R. (1990). *Fitness in special populations*. Champaign, IL: Human Kinetics.
- [29] Statsoft (2004). *Statistica 6*. (1984-2005). Nunspeet, AE: Author.
- [30] Ulrich, D.A. (2000). *Test of Gross Motor Development*. Austin, TX: Pro-ed Publishers.
- [31] Vander Steene, G., & Bos, A. (1973). *Wechsler Preschool and Primary Scale of Intelligence Revised*. Lisse: Swets & Zeitlinger.
- [32] Vermeer, A, & Davis, W.E. (1995). *Physical and motor development in mental retardation*. Basel: Karger AG.
- [33] Wade, M.G. (Ed). (1986). *Motor skill acquisition of the mentally handicapped*. Amsterdam: North-Holland.
- [34] Wechsler, D. (1992). *Wechsler Intelligence Scale for Children* (3rd ed.). Lisse: Swets & Zeitlinger.
- [35] Wheaton, B., Muthén, B., Alwin, D.F., & Summers, G.F. (1977). Assessing reliability and stability in panel models. In D.R., Heise (Ed), *Sociological methodology* (pp. 84-136). San Francisco: Jossey-Bass.
- [36] Woodard, R.J., & Surburg, P.R. (2001). Fundamental gross motor skill performance by girls and boys with learning disabilities. *Perceptual Motor Skills*, 84, 867-870.
- [37] Zittel, L.L. (1994). Gross motor assessment of preschool children with special needs: Instrument selection considerations. *Adapted Physical Activity Quarterly*, 11, 245-260.
- [38] Zittel, L.L., & McCubbin, A.J. (1996). Effect of an integrated physical education setting on motor performance of preschool children with developmental delays. *Adapted Physical Activity Quarterly*, 13, 316-333.

MEASUREMENT OF THE QUALITY OF MATHEMATICS CONCEPTUAL UNDERSTANDING THROUGH ANALYSIS OF COGNITIVE CONFLICT WITH INTERVENTION

Iwan Setiawan HR⁽¹⁾, Ruslan⁽²⁾, Asdar⁽³⁾

Department of Mathematics, Faculty of Mathematics and Natural Science,
State University of Makassar, Makassar, Indonesia.
e-mail address: Iwan.setiawan.hr@gmail.com

Abstract - This research was aimed to obtain information about the quality of mathematics conceptual understanding through analysis of cognitive conflict with intervention. The method used descriptive qualitative. This research was only conducted on students who had cognitive conflict with sample as much as 6 students of class IX consisting of 5 girls and 1 boy students. For collecting data, the instrument used written test and interviews. Each student completed delivering their answer, will be given new information that can lead to cognitive conflict. Based on the results of the research showed that: (1) Cognitive conflict in plane properties occurred in understanding the concept of square with rectangle and concept of rhombus with square (2) Cognitive Conflict for how to find area of plane that happened the concept to find area of parallelogram with rectangle and the concept the area of triangle that had a high line and area of triangle that had high line outside (3) Cognitive conflict happened to the concept of calculating the area of a plane that had been modified (4) Conflict cognitive occurred in understanding the concept of the relationship between area and perimeter of plane, students understand that plane had similar area but different forms will be also had different perimeter.

Keywords: Cognitive Conflict, Intervention, Plane Geometry

I. INTRODUCTION

One of the problems that related to understanding the mathematics as the cognitive processes were contradictions in students' understanding of a particular concept in mathematics. A mathematics concepts can be understood by a variety of information received that related with the concept. This information may include visual representations, examples, or application. If understandings were not mutually integrate into thinking process of students or even contradict each other, it can lead to mental unbalance in his thinking process. Not the integration or contradict understandings these was a condition called as the cognitive conflict. (Ruslan & Asdar, 2011).

So, if there was no common perception (opinion or understanding) which lead to contradiction between two ideas or knowledge that was information / the students' understanding of existing with the new information they will learn, then it was said that had happened conflict between two ideas or knowledge. Like that, if in an individual happen of doubt in choosing one or more than lots of information available in a person it was the conflict. Make a decision or give an answer to specific problems or questions based on specific reasons often create of doubt in the individual to answer it, it was this case we can said had happened conflict in the individual concerned.

Cognitive conflict had become part of the discussion on the theory of psychology, especially in the theory of cognitive development (Cantor, 1983). Piaget (Ernest, 1991) had introduced the concept of cognitive conflict. According to Ernest cognitive conflict occurs when there were conflict between the two schemes of knowledge in the form of cognitive structure inconsistency or contradict each other. The conflict between the two schemes of knowledge according to Ernest can be said as the two schemes of knowledge were not mutually integrate. Thus it can be said that the cognitive conflict was a situation where an individual consciousness encountering imbalance (Mischel, 1971).

For lead to cognitive conflict can be helped by interventions that given the gift of new information related to a concept which understood by students in learning. Giving conflict intervention aims to strengthening the understanding of mathematical concepts students through a guidance or scaffolding in resolving conflicts.

Cognitive conflict in this research was the cognitive conflict between conception namely between concepts/preliminary understanding students and new information as a form intervention. As such cognitive conflict that was used in this research was condition of a child or student had experienced doubts, confusion, did not keep on the establishment, and anxiety during the confronted with giving ideas / new information or by giving the intervention related to the understanding that they can understand / they had before. As well as, Asdar (2012) states cognitive conflict was a conflict in the mind of someone who had comprehensions of concept or application that did not integrate with each other that can be observed in the activity of thinking. In his research also mentions that the cognitive conflict strategy can be used to strengthen geometry conceptual understanding of students in senior high school. This strategy was implemented by giving them information and new knowledge that lead to conflict in understanding of students, then train it to solve the conflict in order to strengthen students conceptual understanding.

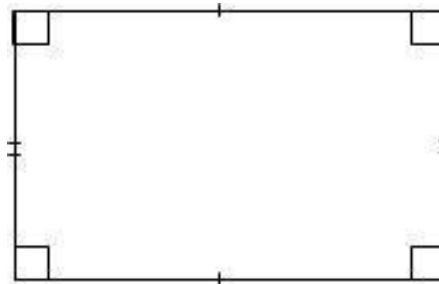


Figure 1. FRectangle

Examples of cases during the observation in school in Makassar, students were shown a that plane figure 1.1 as shown above namely rectangle. Then asked the students to explain what they understand about that plane figure. If the student can to explain the section was rectangle because it had four sides that form a right angle and parallel sides of the same length that means they understand the concept of rectangle. At the time the intervention was given in the form of questions or examples for example, shown square section had properties similar to those mentioned in the rectangle. If the response of the students started to seem doubt, confused or concerned for declare that square was a rectangle, then this condition indicates the students lead cognitive conflict.

Based on the above explanation, researchers want to find out more with the research titled *Measurement of the Quality of Mathematics Conceptual Understanding through Analysis of Cognitive Conflict with Intervention*. With consideration of the subject in this research had studied at the previous level.

II. RESEARCH METHOD

Method of research used descriptive with qualitative approach. This research was only conducted on students who had cognitive conflict with sample as much as 6 students of class IX consisting of 5 girls and 1 boy students. The following scheme subject selection:

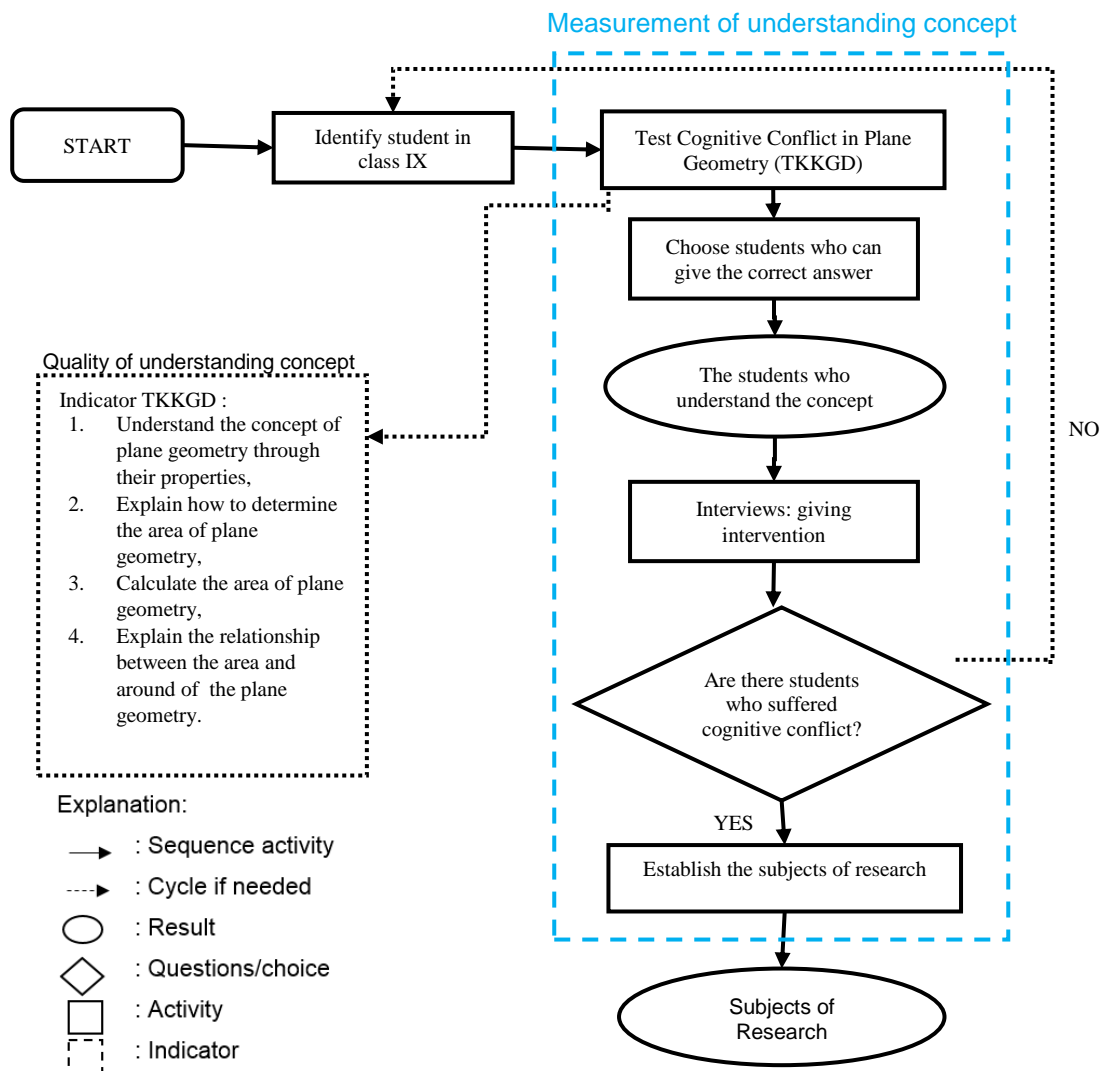


Figure 2. Scheme Selection of the Research Subjects

For collecting data, instrument used written test and interviews. Giving test and interviews in this research was measuring tool to reveal quality of mathematics conceptual understanding. Each student completed delivering their answer, will be given new information (intervention) that can lead to cognitive conflict.

Data analysis in qualitative research, carried out during collecting data in progress, and after the completion of data collection in a particular period. The analysis was performed on each number test about that they did. If the answer that interviewed after the analysis was considered unsatisfactory, the researchers will continue the questions again to a certain extent in order to obtain data that was considered saturated. The data obtained will be analyzed descriptively. The steps in the data analysis, namely: (1) Data reduction, (2) Display Data, and (3) Make conclusion.

III. RESEARCH FINDING AND DISCUSSION

Based on the result of giving test and interviews in school. There are 6 students who have the cognitive conflict, namely: MFA, PAR, MRS, AnP, KhM, and AZA. They experience cognitive conflict in every indicator when it provides intervention. The indicators that are:

- 1) Understanding the concept of plane geometry through their properties
- 2) Describes how to determine the area of a plane geometry
- 3) Calculate the area of a plane geometry
- 4) Explaining the relationship between the area with the around of a plane geometry

For the first and the second indicators each have two questions, and the third and fourth indicators each have one question. So that, the total questions are six questions. Students can be said to in this research experienced cognitive conflict when condition of a child or student have experienced doubts, confusion, does not keep on the establishment, and anxiety during the confronted with giving ideas / new information or by giving the intervention related to the understanding that they can understand / they have before. Where conflict in the minds of students with mental imbalance due to the condition preliminary understanding students of geometry concept/procedure as opposed with new information as a form intervention. So, with intervention for lead to cognitive conflict in students, then this section will be describe the students' cognitive conflict that has been revealed in the results of the data analysis.

In the first section of that is about to understand the concept of a plane through its properties in Question (1) with students who has initial MFA and PAR have experienced cognitive conflict after a given intervention in the form of new ideas. According to their understanding that not all the properties are owned square is also a rectangle. But they run into conflicts when it provides an example of figure where the sample meets the properties of a rectangle but also fulfill properties of a square. They experience confusion, doubt and start inconsistent with their answers. This is indicated by the facial expression suddenly changed as the forehead frown and scratched his head. Their own say "... *Confused* ..." when given the example. They also experienced a contradiction with their understanding after giving about intervention given the length side of the square and rectangular. That beginning they are confident with their answers but suddenly their answers / statement after the intervention different or contradictory to their previous statements.

While on Question (2) with students identified as MRS and ANP also experience cognitive conflict. According to their understanding that the rhombus there that do not have a 90 degree angle so that there are some properties owned by the rhombus is not owned by a square. But they run into conflicts when given a square sample is added diagonal. Here they are confused determine whether the figure is a rhombus because it has a diagonal length of the side the same but all that can be said to be square. Here they seem inconsistent start and they also say "... *Confused* ..." when asked whether the square including a rhombus. They also experienced a contradiction with their understanding after the giving intervention about giving diagonal on the square. That beginning they are confident with their answers and saying that the figure is square, there is based on square owned by the angle of 90 degrees while the rhombus is not all. There also are based on symmetry but suddenly their answers / statement after the intervention different or contradictory to their previous statements.

In the second section that explains how to calculate the area of a plane at question number (3) by students who has initial KhM and AZA have experienced cognitive conflict. According to their understanding that the parallelogram shape different from the rectangular then the formula is different area as well. But they experience cognitive conflict when given examples where the length of the rectangle is equal to the base parallelogram and rectangular width equal to the height of parallelogram. They seem confused and doubt when given the examples and also when asked based on formula is parallelogram include rectangular.

While the question numbers (4) to students identified as MRS and AZA also experience cognitive conflict. According to their understanding that to find the area of a triangle ABC given the high line AD which is inside the triangle. But they experience cognitive conflict when it is located outside the high line and also they are confused and feel doubt to determine the base which is used in calculating the area of a triangle ABC whose height is outside. And they are also confused whether the area of a triangle whose height is equal to the inside if height is outside the triangle ABC.

In the third section which is about calculating the area of plane in question numbers (5) with students who has initial KhM and ANP also experience cognitive conflict. According to their understanding in this question to calculate the area of the shaded area that combines all the shaded area thus forming a half of a square, it could be a way of breaking down the last part in the wide count each part. But when they experience cognitive conflict in the wake of the change into a new form, but remained virtually identical form, it's just been modified. Because when given the same hatching with previous figure is puzzled why they appear in the shaded area are the same but different shape.

In the fourth section which describes the relationship between the area with the circumference of a plane at question numbers (6) with students who has initial MFA and PAR also experience

cognitive conflict. According to their understanding though the same area if the size is different than around will be different. But they experience cognitive conflict when determining if the area around largest square length is 100cm^2 . They just think that the flat side of a wake up only limited integers only. So they seem free to determine the greatest around that when unlimited.

So for the quality of students' understanding the based on cognitive conflict analysis that described above can be said as follows:

- 1) For students who experience cognitive conflict that *Not Consistent* can be said an understanding of the students are still less because students do not know the correct answer and still do not have a curiosity about the correct answer, because it is based on the answers they explain from the beginning contradict with explaining after giving the intervention.
- 2) For students who experience cognitive conflict that feel *Confused* can be said understanding of the students are enough because the students simply cannot distinguish the correct answer but the curiosity to know the correct answer was there.
- 3) For students who experience cognitive conflict that feel *Doubt* can be said an understanding of the students are good enough because the students know the correct answer, but rather difficult to explain the reason.

IV. CONCLUSIONS AND SUGGESTION

From the description above, it can be concluded that the description: (1) Cognitive conflict in plane properties occurred in understanding the concept of square with rectangle and concept of rhombus with square (2) Cognitive Conflict for how to find area of plane that happened the concept to find area of parallelogram with rectangle and the concept the area of triangle that had a high line and area of triangle that had high line outside (3) Cognitive conflict happened to the concept of calculating the area of a plane that had been modified (4) Conflict cognitive occurred in understanding the concept of the relationship between area and perimeter of plane, students understand that plane had similar area but different forms will be also had different perimeter.

For the quality of student understanding based cognitive conflict can be categorized, that: (1) For students who experience cognitive conflict that *Not Consistent* can be said an understanding of the students are still less. (2) For students who experience cognitive conflict that feel *Confused* can be said understanding of the students are enough (3) For students who experience cognitive conflict that feel *Doubt* can be said an understanding of the students are good enough.

Suggestions for improvement of learning geometry, especially in the plane geometry:

- 1) Results from the research showed students as subjects of research experienced cognitive conflict in understanding the concept of plane through their properties. This cognitive conflict happen because understanding the students in understanding the concept of plane through their properties not mutually integrate well with their understanding. This is very important because the most basic thing to understand the concept of plane. When the student is able to understand the concept of plane through their properties so they are able to differentiate between plane and with each other. And also can understand the relationship between plane with each other with a formula approach or properties owned by each plane.
- 2) Material of the lesson about plane should be using differentiation for see the relationship between plane and the other plane. In fact, if the terms of the properties of each plane have correlation between plane one another especially rectangular. Therefore, it is suggested in the learning that integrate concepts between one plane and with each other.

REFERENCES

- [1] Asdar. 2012. Pemanfaatan Strategi Konflik Kognitif Dalam Pemecahan Masalah Matematika Untuk Memperkuat Pemahaman Konsep Geometri Siswa SMA di Kota Makassar. *Research Report*. Universitas Negeri Makassar. (references)
- [2] Cantor, G. N. 1983. Conflict, learning, and Piaget: comments on Zimmerman and Blom's "Toward an empirical test of the role of cognitive conflict in learning". *Developmental Review*. 3, 39-53.
- [3] Piaget, J. 1985. *The Equilibration of Cognitive Structure: The Central Problem of Intellectual Development*. The University of Chicago Press, Chicago.

- [4] Mischel, T. 1971. Piaget: *Cognitive Conflict and Motivation of Thought, Cognitive Development and Epistimology*, 21. 265-287
- [5] Ruslan & Asdar. 2011. Potential of Cognitive Conflict in Students Understanding About the Limit Function. *Research Report*. Department of Mathematics, Faculty of Mathematics and Natural Science, UNM Makassar.
- [6] Ruslan & Asdar. 2012. Measurement of Potential of Cognitive Conflict in Students Understanding About the Limit Function. "Proceding Koferensi Ilmiah Nasional HEPI" *Assesment dan Pembangunan Bangsa Surabaya*, 14-15 December 2012. UNS Page 328-333. ISBN 978-979-028-526-2. (references)

MODIFICATION OF RANDOMIZED ITEMS SELECTION AND STEP-SIZE BASED ON TIME RESPONSE MODEL TO REDUCE ITEM EXPOSURE LEVEL OF CONVENTIONAL COMPUTERIZED ADAPTIVE TESTING

Iwan Suhardi

Department of Electrical Engineering, Faculty of Engineering
State University of Makassar, Makassar, Indonesia
email: iwansuhardi@yahoo.com

Abstract - To estimate the ability of students, computerized adaptive test (CAT) has been shown to be more efficient than conventional tests using paper and pencil test (P&P Test) and computerized-based testing (CBT). However, the conventional CAT applying the Maximum Likelihood Estimation (MLE) and the step-size to estimate the ability of the test taker and the application of item information function (IIF) for the selection of items displayed will cause item exposure problems or frequent emergence of some items that given to test participants, so that the item was easy to spot, especially at the beginning of the emergence in the order items. This paper attempts to determine the effect modification by randomization in the CAT algorithm and step-size based on the response time to estimate the ability of the test taker. Items bank for the research using item response theory models one parameter logistic (1 PL). Development model is the method of randomization using 5-4-3-2-1 models based on MLE and grouping the response time for a constant step-sizenya. Based on study results, the CAT algorithm modification resulted in the appearance items are more varied, thereby reducing exposure item problem without reducing the efficiency of CAT.

Keywords: *computerized adaptive test, item exposure, randomization, step-size*

I. INTRODUCTION

Generally, the test was built to estimate the ability of participants test. Giving the test too easy for the person taking the test is a waste of time and otherwise, the questions that are too difficult, also produces test scores that are not informative. For customizing a test to bring the level of ability of each individual participant test, a solution should be sought. In the development of modern measurement theory, item response theory (IRT) as well as the advances in computer technology makes it possible to develop a computerized adaptive test, or more popularly known as the Computerized Adaptive Testing (CAT) [1] [2]. Known as "computerized", because the implementation of computerized testing really is no longer using paper and pencil. "Adaptive", because items have been selected based on the results of the self-regulatory analysis and adapted to the needs or abilities of the examinees, works automatically through a computer software. According to Weiner, CAT is a test held for participants where the items are determined based on the response from the participants' answer [2].

Comparison between traditional testing and adaptive testing in Table 1 below [3]:

Table 1. Comparison Between Traditional And Adaptive Testing

FACTOR	TRADITIONAL TESTING	ADAPTIVE TESTING
Composition Test	Each participant received a similar test	Each participant received a different test
Difficulties Test	Intended for the average participant	Intended for individual participants
Length Test	Identical for all participants of the test. In general, a long test	The length of the test is not the same for all participants. In general, a short test
Time Test	Certain time	Anytime
Organization Test	It takes a long time	It takes a short time
Results Instantly	Generally, it requires a long time to see results	The results appear instantly

Efficiency of CAT is supported by several studies. McBride and Martin concluded that to achieve the same level of reliability, the conventional tests still require as much as 2.57 times the number of items in adaptive test [4]. Similar research results by Eignor also concluded that the adaptive test only requires a long test less than half of the length of the paper and pencil test at the same level of precision measurements [5].

However, application of the maximum likelihood method and step-size at CAT for estimating the ability of test takers have a number of items given to participants of the test appear more often compared other items. This occurs especially at the beginning of the emergence of items given to participants of the test. Therefore, it is necessary to modify the CAT algorithm to reduce the problem of often appear items and are easily recognizable. This is known as item exposure. Although the design of adaptive test is more efficient and reliable, it is not guaranteed as safety testing because of frequent emergence of certain items.

II. THEORY OF CAT

The practical steps commonly used to develop conventional CAT algorithm are as follows [6]:

A. CAT Starting Point

If no preliminary information about the ability of participants, the CAT can begin by selecting the items beginning with a medium level of difficulty [7] [8].

B. Continuing Process

After obtaining the response of the participants' answers, CAT system gives a response assessment with a correct or incorrect answer. There are two steps to continue the process of CAT, which estimate the ability of the participants and how to choose the next items.

1. Method of Estimating Capabilities

Having answered the first item given, the ability of the test taker is estimated based on the parameters of items, the estimated value of the initial capabilities, and answers of the items whether true or false. The general method used to estimate the ability of the test taker is Maximum Likelihood Estimation (MLE) [9] [10]. One problem with the application of MLE method on adaptive testing is the inability of the likelihood function to find a solution when there is a maximum of examinees who earn a score of 0 (answered incorrectly on all items) or a perfect score (answered correctly on all items). To overcome the problem of the inability of MLE method in estimating the level of response capability when the participants have not figured test participants can use the method step size [11] [12]. Based on the method step size, ability level test participants increased or decreased by a certain number during the test have not been patterned response.

2. Selection of Next Items

Once the ability of participants is estimated, the computer select the next items. Lord suggests using items maximum information procedure to select the next items to be given to the participants of the test [13]. Based on this procedure, item that has a greatest value information function on the ability of certain participants have to be given to the test taker. This ensures that the value of the function test information for each person taking the test is maximum, meaning that the standard error of measurement (SEM) minimum because no other test information function is the inverse of the variance of the measurement error. In other words, this method

guarantees yield prediction skill level of participants with high accuracy [5]. Using the information function, the accuracy of measurement in estimating the ability of test takers can be calculated at every level of ability. Function Birnbaum information item to be stated by the following equation [14]:

$$I(\theta) = \frac{2.89 a_i^2 (1-c_i)}{[c_i + \exp(1.7 a_i(\theta - b_i))] [1 + \exp(-1.7 a_i(\theta - b_i))]^2} \quad (1)$$

The above equation shows that the information is only dependent on item parameter (eg a, b, and c for the model 3P) and the level of ability (θ). Test information function is the number of item information function test developers [15]. Information test function device is mathematically written as follows:

$$TIF = \sum_{i=1}^n I_i \quad (2)$$

As an item information function, the test information function illustrate how accurate the estimate test for different ability levels. The greater level of information on given ability, the more accurate the estimated ability of the test device. Standard error of measurement (SEM) is expressed by the following equation [15]:

$$SEM = 1/\sqrt{TIF} \quad (3)$$

C. Stopping Rule

Two main methods are used to stopping CAT, equal measurement precision and fixed number item. Both of these methods produce different measurement error variance. The purpose of the equal measurement precision method is generating test scores with the same error rate measurements for each test taker's ability. Standard error of measurement equivalent set a limit on 0.30 with a reliability of 91% on conventional tests [16]. But in practice it is also use criteria fixed number of items, the dismissal rules CAT, eg using criteria fixed starting rule as much as 20 items to avoid the process of tests that may not converge.

In this study, two draft adaptive test developed are a conventional CAT (not randomized) and a randomized. The design of a randomized CAT is principally the same as the design of conventional adaptive tests. The difference lies in the selection of items for second item and subsequent use of randomization principle 5 – 4 – 3 – 2 – 1. To estimate the ability of the test taker, when the response of participants has not been patterned, used the method step-size based on the response time.

Participants test that have a high skill level is assumed to be able to answer the item correctly in a faster time than the learners who have less ability levels. Use of the speed test participants' responses factor to the additional assessment information is also recommended by Dunkel [17]. Van der Linden said that if the speed of response and accuracy-related or if both are important in the context of the test, the speed of response can be included in the assessment rubric [18]. Lidia Martinez' research on CBT found that the group spend an average time to respond to the fastest initial test items have an average score higher. However high or low the average score is statistically not influenced by the length of time the person taking the test to review the previous item [19]. This indicates that the speed in responding to the items correctly influenced by the ability of the test taker.

The results of the research are almost the same also delivered by Phil Higgins. Group of test participants with high scores able to complete the items properly with the average time that is shorter than the test group of participants with moderate and low scores [20]. In another study on the CBT, Higgins also found that higher levels of item difficulty, the person taking the test will need more time to answer and review the item [21]. This shows that the response time test takers work item correctly correlated with the level of ability of the test taker.

Chang's research concluded that there was no statistical difference with regard to gender and origin to the test [22]. Therefore, the additional variable response times on the step size method can be applied to all the participants of the test without implications for gender and origin to the test.

III. RESEARCH METHOD

Items bank for the purposes of this research consisted of 600 items based on IRT 1 PL models. In this model, chances for somebody answered the item correctly depend only on the parameter level of difficulty items. Furthermore, two draft adaptive test developed is a conventional CAT (not randomized) and randomized. In this study the method of estimating the ability of test takers using MLE method and step-size.

In conventional CAT, the method of selecting the first item using medium difficulty level that starts with a range between -0.5 to 0.5 were selected randomly. Ability level estimation method using maximum likelihood estimation, but when the response answers the test taker is not yet patterned, estimating the level of ability of using a step size of 0.5. The next method of selecting items using the criteria of maximum information function. Items that have greatest value information function on specific capabilities have to be given to the test taker.

In the design of a randomized CAT, the design principle is the same with the conventional adaptive tests. The difference, at the election of the second point and so on using the principle of 5 – 4 – 3 – 2 – 1. The second item been selected randomly out of five (5) items which have the greatest information functions, the third item been selected randomly out of four (4) items which have the greatest information functions. The fourth item been selected randomly out of three (3) items which have the greatest information functions. The fifth item been selected randomly out of two (2) items which have the greatest information functions. Furthermore, for the sixth items and so the criteria for selecting the next items back to the maximum information function criteria that are not randomized (1 item).

To estimate the ability of the participants during the response of participants that not yet patterned, hence used the method step-size with an additional variable response time. For example, if there is no further information about the prior ability level of participants test, so the value $\theta_0 = 0$. Interval step size steadily increased of k (in this study was taken the value of $k = 0.5$). If the test participants responded with incorrectly answer, then estimate the ability of the participants into $\theta_0 - k$ or $0 - 0.5 = -0.5$. Meanwhile, when the the participants answered correctly, the estimated ability of the participants becomes $\theta_0 + x k$, or 0.5. x , where x is a positive constant multiplier and the amount depends on the category of the response time when the participants answered correctly. The procedure for estimating the level of participants ability with a step-size by a factor of response time participants are shown in Table 2.

Table 2. Estimated of Participants Ability Level in Step-Size Method Based on Response Time

Specification Notation: time limit = 150 second θ_0 = basic ability level = 0 k = step size = 0,5 x = constant multiplier $\theta_{k+i} = \theta_{i-1} + xk$ (for the correct response) $= \theta_{i-1} - k$ (for the incorrect response)		Response Corect Answer Consecutively				Response Incorect Answer Consecutively			
		1 st item	2 nd item	3 rd item		1 st item	2 nd item	3 rd item	
		θ_1	θ_2	θ_3		θ_1	θ_2	θ_3	
Response Time Category	Very Fast: $x = 1.8$ (Less than 30 s)	0,9	1,8	2,7		-0,5	-1,0	-1,5	
	Fast: $x = 1,6$ (31 to 60 s)	0,8	1,6	2,4		-0,5	-1,0	-1,5	
	Medium: $x = 1,4$ (61 to 90 s)	0,7	1,4	2,1		-0,5	-1,0	-1,5	
	Slow: $x = 1,2$ (91 s.d. 120 s)	0,6	1,2	1,8		-0,5	-1,0	-1,5	
	Very Slow: $x = 1$ (121 to 150 s)	0,5	1,0	1,5		-0,5	-1,0	-1,5	
Extra time for 150 seconds. More than 300 s, the response is considered incorrectly.		0,5	1,0	1,5		-0,5	-1,0	-1,5	

In this study, the test termination criteria used were the test is stopped if the estimated value of the SEM has reached 0.30.

IV. RESULTS AND DISCUSSIONS

Summary of items bank statistics used in this study as follows:

Table 3. Statistics of Items Bank

General Description	Based on IRT with 1 PL
	The number of items = 600 items
Difficulty Level (b)	Minimum value = -3.0
	Maximum value = 3,0
	Number of items with medium difficulty level = 101 items

CAT test results showed that the number of items with a medium difficulty level, between -0.5 to +0.5 totaled 101 items. This case means that the possibility of the first items that appear to have the possibility a number of 101 items were taken randomly. When the participants answered correctly, then the second item to be displayed is item with the maximum information for $\theta = 0.5$, and when the participants answered the item incorrectly, then the second item were to be displayed is items with the maximum information for $\theta = -0.5$. So it can be ensured that the conventional CAT, the second item consists only of the possibility one of the two items. In this study, the second item that appear are No.ID 577 (if answered correct) and No.ID 405 (if the answer is incorrect). Often the appearance of numbers No.ID 577 and No.ID 405 made a test CAT become unsafe due to the familiar questions.

The other case of item exposure that often appear is when using the step-size method. If the participants answered the questions always correct then the items appear are the items that have the maximum information value for $\theta = 0.5, 1.0, 1.5, 2.0$, and 2.5 , ie the second item with No.ID 557, the third item with No.ID 121, the fourth item four with No.ID 105, the fifth item with No.ID 247, and the sixth item with with No.ID 255. But, if the participants answer is always incorrect, the items appear is a matter that has maximum information value for $\theta = -0.5, -1.0, -1.5, -2.0$, and -2.5 , ie the second item with No.ID 557, the third item with No.ID 405, the fourth item with No.ID 125, the fifth item with No.ID 204, and the sixth items with No.ID 129.

If the participant's answer responses has been patterned (response answers already are correct and incorrect answers), then the next appeared item was quite varied because the first items that appears has a variable items is relatively large (101 items). However, with the use of maximum information function model to find items that match the level of the test participants' ability estimation, it is possible a lot of items that can not be displayed because they never get the maximum value for all levels of ability.

One proposed solution is to use a step-size method is based on the response time of participants answered correctly. Response time were stratified into groups based on the speed of response of participants correctly answered items raised by CAT. In the method step-size based on the response time, the formulation of the magnitude of the step-size value given additional constant multiplier based on the response time. The faster students respond to answers correctly then the bigger the multiplier constants.

Another proposed solution is to randomizes the maximum value of the function information. When the conventional CAT determines the items appear based on the maximum value of the function information (single), the CAT model of randomisation determines items appear by randomizing the maximum information function based group $5 - 4 - 3 - 2 - 1$.

The second items obtained from randomize 5 greatest value function information based on the premise that in the early stages of the estimated level of proficiency test participants still contains an error value (SEM) high, so that not affect the result estimates the level of ability of the participants. Along with the many steps to estimate the ability of participants, the group randomized increasingly scaled down (to $4 - 3 - 2 - 1$) along with decreasing error estimate. Thus, the items appear still refer to the estimated rate of the test participants' ability and does not affect the length of the test.

As an example will be given some comparative results of conventional CAT (which is not randomized) and the randomization CAT, as follows:

Table 4. Example of Conventional Cat (Not Randomized)

ITEM	1 st	2 nd	3 th	4 th	5 th	6 th	7 th	etc
No. ID	284	577	121	105	247	430	283	etc
Response	Correct	Correct	Correct	Correct	Incorrect	Correct		etc
No. ID	25	577	121	105	357	126	92	etc
Response	Correct	Correct	Correct	Incorrect	Incorrect	Correct		etc
No. ID	146	577	121	518	139	450	77	etc
Response	Correct	Correct	Incorrect	Incorrect	Incorrect	Correct		etc

Table 5. Examples Of Randomization Cat

Item	NO.ID	Response Answer			Explanation		
1	I.003 (b = 0.1)	Correct			First item is selected based on the level of medium difficulty (- 0.5 ≤ b ≤ 0.5)		
Selection Process of Second Item							
Because CAT is not patterned, then the process of selecting second items, using the method of step-size							
value of Θ that may appear	Θ Selected	Alternative 5 Greatest Value of IIF					
		IIF	No. ID	No. ID Selected	No.ID Selected appears by selecting randomly from 5 greatest value of IIF		
		0, 9	0.7	0,7225		I-068 (b = 0.7)	I-170
		0,8		0,722447802		I-170 (b = 0.69)	
		0.7		0,722447802		I-480 (b = 0.71)	
		0,6		0,722291238		I-034 (b = 0.68)	
0,5	0,722291238	I-406 (b = 0.72)					
Item	NO.ID	Response Answer			Explanation		
2	I-170 (b = 0.69)	Correct			Second Item is selected based on the results of randomization		
Selection Process of Third Item							
value of Θ that may appear	Θ Selected	Alternative 4 Greatest Value of IIF					
		IIF	No. ID	No. ID Selected	No.ID Selected appears by selecting randomly from 4 greatest value of IIF		
		1.6	1.2	0,7225		I-156 (b = 1.2)	I-122
		1.5		0,722447802		I-407 (b = 1.19)	
		1.4		0,722447802		I-122 (b = 1.21)	
		1,3		0,722291238		i-215 (b = 1.18)	
1.2							
Item	NO.ID	Response Answer			Explanation		
3	I-122 (b = 1.21)	Correct			Third Item is selected based on the results of randomization		
Selection Process of Fourth Item							
value of Θ that may appear	Θ Selected	Alternative 4 Greatest Value of IIF					
		IIF	No. ID	No. ID Selected	No.ID Selected appears by selecting randomly from 3		

2.1	1.7	0,7225	I-167 (b = 1,70)	I-085	greatest value of IIF	
2.0		0,722447802	I-085 (b = 1,69)			
1.9		0,722447802	I-499 (b = 1,71)			
1,8						
1.7						
Item	NO.ID	Response Answer		Explanation		
4	I-085 (b = 1,69)	Incorrect		Fourth Item is selected based on the results of randomization		
Selection Process of Fifth Item						
Because CAT is already patterned, then the process of selecting fifth item, using the method of MLE						
Value of Θ	IIF	No. ID		N0. ID Selected		
1.71	0,72250000	I.499 (b = 1.71)		I.165		
	0,72244780	I.165 (b = 1.70)				
Item	NO.ID	Response Answer		Explanation		
5	I-065 (b = 1,70)	Incorrect		Fifth item selected appears by selecting randomly from 2 greatest value of IIF		
Selection Process of Sixth Item						
Value of Θ	IIF	No. ID		N0. ID Selected		
1.40	0,72250000	I.279 (b = 1.40)		I.279		
Selection Process Sixth item and Subsequent : The sixth item and subsequent items are selected based on the value of the largest IIF						

From the results of Table 3, it looks that CAT were not randomized appeared several items with the same identity, especially in the earlier pattern of the use of CAT. Meanwhile, in Table 4, the randomized CAT, many variations of possibilities items appear although the participant answers the same pattern. With so many variations items appear in the CAT randomized, it can reduce the level of item exposure, so that would make CAT more secure. Variations items appear on the actual randomized CAT has a difficulty level that is not much different from the CAT that were not randomized so it will not affect the length of the CAT test.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

From the analysis and discussion above, it can be concluded that the CAT by the method of randomization maximum information function with the criteria 5-4 - 3 - 2-1 by applying Maximum Likelihood Estimation (MLE) and the step-size based on the response time to estimate the ability of participants can bring items with more variety. On the same participant answer pattern, the randomized CAT has item variation that have difficulty level similar to non-randomized CAT. Thus, it would produce a reduction in the level of exposure to the CAT items and increase security without reducing the CAT efficiency.

B. Recommendations

From this study, the design of randomized algorithms CAT is recommended to be applied to the adaptive test algorithms. CAT randomized algorithm does not reduce the level of efficiency and precision measurement, items on the initial order granted to the participants the test more varied so as to improve the safety test. This study uses a model 1 PL, it is recommended to use the model 2PL or 3PL to better examine variations items appear in the CAT.

REFERENCES

- [1] F.M. Lord, "Applications of item response theory to practical testing problems", Hillsdale, NJ : Lawrence Erlbaum Associates. 1980.

-
- [2] H. Wainer, "Computerized adaptive testing: A primer, " 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates, 1990
 - [3] J.Q. Tian, D.M. Miao and X. Zhu, "An Introduction to the Computerized Adaptive Testing. US-China Education Review", 2007,1, pp. 72-81.
 - [4] J.R. McBride and J.T. Martin, "Reliability and validity of adaptive ability tests in a military setting". in D.J. Weiss, (Ed), New Horizons in Testing, New York, NY: Academic Press,1983, pp.223 – 236.
 - [5] D.R. Eignor, M.L. Stocking, M.L. and W.D. Way, " Case studies in computer adaptive test design through simulation", Research Report, Princeton, NJ: Educational Testing Service, 1993, pp. 93 – 96.
 - [6] D. Thissen and R.J. Mislevy, "Testing algorithms", dalam H. Wainer (Ed.), Computerized Adaptive testing : A Primer, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates, 1990, pp. 103-135.
 - [7] W.P. Vispoel, " Creating computerized adaptive test of music aptitude: Problem, solutions, and future directions," in F. Drasgow and J.B. Olson-Buchanan (Eds.), Innovation in Computerized Assessment Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 1999, pp. 151-176.
 - [8] C.N. Mills, "Development and introduction of a computer adaptive graduate record examination general test," in F. Drasgow & J.B. Olson-Buchanan (Eds.), Innovation in Computerized assessment, Mahwah, NJ: Lawrence Erlbaum Associates Publishers. Mills, 1999, pp. 123.
 - [9] A. Birnbaum, "Some latent trait models and their use in inferring an examinee's ability," MA: Addison-Wesley, 1986.
 - [10] F.B. Baker, " Item response theory: Parameter estimation techniques," New York: Marcel Dekker, Inc., 1992.
 - [11] B.G. Dodd, "The effect of item selection procedure and step-size on computerized adaptive attitude measurement using the rating scale model," Applied Psychological Measurement, 4, 1990, pp. 355 – 366.
 - [12] D.J. Weiss, "Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education," Measurement and Evaluation in Counseling and Development, 37, 2004, pp. 70-84.
 - [13] F.M. Lord, "A broad-range tailored test of verbal ability," Applied Psychological Measurement, 1, 1977, pp. 95-100.
 - [14] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, "Fundamentals of item response theory," Newbury Park, CA: Sage Publication, Inc., 1991.
 - [15] K. Hambleton and H. Swaminathan, " Item response theory," Boston, MA: Kluwer Inc., 1985, pp. 94.
 - [16] D. Thissen, "Reliability and measurement precision," in H. Wainer (Ed.), Computerized adaptive testing: A Primer (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates, 1990, pp. 103-135.
 - [17] P.A. Dunkel, "Considerations in developing and using computer adaptive tests to asses second language proficiency" [Version Electronic]. Washington DC : Eric Clearinghouse On Languages And Linguistics Center For Applied Linguistics, 1999, <http://www.cal.org/resources/digest/cat.html>.
 - [18] W.J. Van der Linden, D.J. Scram and D.L. Schnipke, "Using response-time constraints in item selection to control for differential speededness in computerized adaptive testing" [Elektronik Version], LSAC Research Report Series, Newton, PA Law School Admission Council, 2003, <http://www.lsac.org/lisacresources/Research/CT/CT-98-03.pdf> .
 - [19] Lidia Martinez, "Time usage and candidate performance" [Electronic version], Chicago: Measurement Research Associates, Inc., 2009, <http://www.rasch.org/mra/mra-06-09.htm>
 - [20] Phil Higgins, "Candidate measured ability and use of time", [Elektronik version], Chicago: Measurement Research Associates, Inc. 2009, <http://www.rasch.org/mra/mra-10-09.htm>.
 - [21] Phil Higgins, "Item difficulty and time usage", [Elektronik Version], Chicago: Measurement Research Associates, Inc., 2009, <http://www.rasch.org/mra/mra-05-09.htm>
 - [22] S.R. Chang, B.S. Plake and A.A. Ferdous, "Response times for correct and incorrect item responses on computerized adaptive testing", 2005, Paper presented at the 2005 annual meeting of the America Education Research Association (AERA), Montreal, Canada.

CHARACTERICS OF AN INSTRUMENT OF VOCATIONAL INTEREST SCALES^{1,2}

Kumaidi

The Muhammadiyah University of Surakarta
kuma_426@yahoo.com

Abstract - This paper presents part of study results conducted in 2014. The original study was aimed at developing An instrument of vocational interest scales to guide students in selecting specific career pathways that meet his/her personalities. The instrument developed based on self-directed search from Holland (1985) and it was modified to make it suitable with Indonesian students. The item scales were analyzed and judged by six psychologists work at the Psychology Department of The Muhammadiyah University of Surakarta. It was found that 216 out of 276 items have at least 0,61 CVI (Aiken, 1985). When the scales were tested to those students at various job lines, it was found that the scales quite well differentiated students of having different group of vocational track of occupational pathway. The study concluded that the scales were worthwhile to be used as an instrument to guide Junior High School leavers in selecting departments and/or specialist pathways in Vocational High Schools.

Keywords: *vocational interest scales, validity of the scales, person-environment interaction*

I. INTRODUCTION

Many young children who is just graduated from Junior High Schools do not know where should go when they wish to continue their schools. Many of them especially those students chose a vocational track of occupational pathway in a Vocational High School without appropriate guidance. They chose a vocational track of occupational pathway almost randomly, following where their peer group had gone to a vocational school. Therefore, it was thought that an instrument of vocational interests scales would be useful to make direction of those students.

The study reported here originally was conducted in order to develop an instrument of vocational interest scales which might be useful to guide those students in selecting a career pathway, by choosing an appropriate department or a vocational track of occupational pathways. A career pathway is a vocational track of occupational pathways offered by a Vocational High School which prepares its graduate to persue occupations or jobs as intended by their training. For example, students who wish to be motor cycle mechanics they may attend to the "Technical Motor Cycle Training Program" in a Vocational High School.

In developing such scales, researcher selected Self Directed Search (SDS [1]) as the basic instrument and decided to translate and modify the items where needed. The SDS was prepared based on personality theory called person-environment interaction [1]. The theory states that every people has one's own type of personalities and every occupation (or a job) also has its own type of personalities. The type of personalities in every occupation (job) is developed based on the types of people who are supported the occupation (job). A specific occupation generally attracts people with similar types of personalities. Therefore, an occupation or a job is declared to be having certain type of personalities due to people in the occupation will develop their own work environment.

¹ The author of this study express his gratitude to the Directorate of Technical and Vocational Education of the Ministry of Education and Culture of the Republic of Indonesia for their funding of the study

² Paper presented at the Interntional Conference on Educational Research and Evaluation, held at Yogyakarta State University 30-31 May 2016

According to [1], an individual person is nurtured in one's social environment throughout one's life. The environment started from the family, neighbors, schools, and peer groups. One's personalities were developed and nurtured as results from the interaction of one's perceptions, self-concepts, values, achievement, performance, reaction to the environment challenges, and so on. One might make selection to what make one preferences, likes, and dislikes to certain types of those opportunities given by one experience throughout one life.

Every person might be characterized as one of six types of personalities that are realistic, investigative, artistic, social, enterprising, and conventional which is popularized by its abbreviation as RIASEC [1]. The environment types are also identified with similar characteristics as R-typed of environment, and so on. People with R-typed of personalities intended to group and develop R-typed of environment. People with R-typed of personalities tend to have preferences in realistic occupations such as an electrician, a mechanical-engineering technician, a mechanic or repair worker to say a few. The other types of personalities have their own types of preferences to types of occupations.

People with certain types of personalities possesses certain preferences, competencies, self-perceptions, values, and other characteristics which may differentiate from other people with different type of personalities. This typological of personalities concerning person and environment might be used to develop an instrument to tap and to match the similarity between person and suitable environment. The suitable environment could be schooling such as a vocational track of occupational pathways offered by a Vocational High School or work environment. This basic concept could be used as the basic theory to develop an instrument of vocational interest scales.

An instrument of vocational interest scales may be used to identify pupils type of personalities. The instrument of vocational interest scales may also be used to identify occupations' characteristics in term of its vocational personalities based on its work environment. In turn, school counselors may use this instrument to guide pupils in selecting a training program that has similar types of personalities. It is hoped, by matching between pupils and training program characteristics of personalities, student achievement and performance in schools might be improved.

II. METHOD OF STUDY IN DEVELOPING THE INSTRUMENT

There are many instruments of vocational interest scales. One of them is Self Directed Search (SDS) published in the book by Holland [1]. This SDS was modified by Aljufri and Kumaidi (1990, [3]). One of the reasons in choosing SDS over the others is the fact that using SDS to guide pupils future schooling in a career track typed schools does not need advance training in psychology.

Following earlier work by [3], this development of an instrument of vocational interest scales was prepared. The modification made by inserting/changing some types of activities, competencies, and occupations tends to be perceived to be preferred by Indonesian pupils. The initial draft of the instrument then was reviewed by six psychologists from the Faculty of Psychology of the Muhammadiyah University of Surakarta in order to validate the appropriateness of the items to the corresponding indicators (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985 [4]; Lawshe, 1975 [5]).

After the responds from the psychologists were collected, an index of item validity was calculated. The index of item validity is suggested by Aiken [2]. The instrument of vocational interest scales was then tested to 1329 third year students enrolled in 43 out of 128 vocational tracks of occupational pathways in Senior High Schools in the provinces of Daerah Istimewa Yogyakarta (DIY) and Jawa Tengah (Jateng). An analysis of construct validation was performed using LISREL 8.80 [6].

The final analysis was directed to identify whether the instrument could differentiate students from different vocational tracks of occupational pathways in Senior High Schools. This analysis actually was a descriptive statistic one. Students having typological preferences in certain occupations as they enrolled in different vocational tracks of occupational pathways in Senior High Schools were identified based on their three highest scores of the six types.

III. RESULTS AND DISCUSSIONS

Results for the review process to evaluate the appropriateness of items to measure the indicator developed for the instrument of vocational scales as follows. This review is thought to evaluate the score interpretation based on content related validation, as defined by [5]. From the total of 276 items, 216 items were selected with at least having a content validity index (CVI) [2] = 0,61. These items were considered valid items. Therefore, the instrument of vocational interest scales in this study consists of 216 items having at least 0,61 index of content validity.

Summary analysis using LISREL 8.80 to evaluate construct validity of the instrument is presented in Table 1. From Table 1, it may be inferred that items' fitness to the attribute measured is relatively good, although it is need improvement in the next step of development. It needs further revision on the selected items to make it closer to the defined construct of each aspect.

Table 1. The Fitness of Costruct Measured by the Instrument of Vocational Interest Scales

Aspects	p	RSMEA	Interpretation
Realistics (R)	0,057	0,05	Fit
Investigative (I)	0,052	0,05	Fit
Artistic (A)	0,057	0,01	Fit
Social (S)	0,046	0,05	Fit enough
Enterprising (E)	0,051	0,04	Fit
Conventional (C)	0,051	0,05	Fit

The scoring system applied to identify students typological personalities is based on one-zero (1-0) scoring system. The score one is given to students who are agree, preferred, or likes to statement of an item. Their declaration indicated that they are preferred to activities, like to occupations, or possesses compencies stated in the items therefore to the item is given score one (1), otherwise is given score zero (0). The reliability coefficient for scores of the six scales which is estimated using Cronbach Alpha can be seen in Table 2. From Table 2, the score on the Realistic Scale which has coefficient of Cronbach Alpha about 0,91 is the largest among the six, while score of the Social Scale which has coefficient Cronbach Alpha around 0,85 is the lowest. It is concluded that the scores assigned to students for each scale are judged as highly reliable.

Table 2. Reliability Coefficient Estimates of Cronbach Alpha for the Vocational Interest Scales

Scales	Cronbach Alpha	Interpretation
Realistics (R)	0,91	reliable
Investigative (I)	0,89	reliable
Artistic (A)	0,87	reliable
Social (S)	0,85	reliable
Enterprising (E)	0,86	reliable
Conventional (C)	0,89	reliable

The next analysis was performed to identify the typological personalities of students and school programs that offer certain vocational track of occupation pathways. [1] suggests that combination of two or three aspects (scales) of RIASEC is more stable identification of one's work environment than the single aspect. Therefore in this study, researcher decided to use combination of three aspects (score scales) of the RIASEC.

Those students who have highest scores on the first three highest of the six, will be identified as persons with personalities of the three highest. For instance, one may be identified as an artistic person due to he/she scored highest on scales A, S, and E. This person would be classified as having personalities of ASE. Identically, for school programs with certain vocational track of occupation offered by a High Vocational School will be identified as ASE if the programs attended by a group of students with typological ASE.

Table 3. Personality Types of the 10 School Programs in High Vocational Schools*)

No.	School Programs in the track of Occupations	Personality Types
1.	Skin Counselors	SEA
2.	Dress Designer	ASE
3.	Concrete Workers	RES
4.	Hotel Acomodation Workers	AES
5.	Wood Construction	RSE
6.	Book Keeper	SCA
7.	Workers in Agro Industry	CSE
8.	Visual-Communication Designer	RSE
9.	Metal Art Designer	AES
10.	Office Management	SRE

*) The study completed 43 out of 128 school programs in High Vocational Schools

Applying that system of scoring to identify person and school program type of personality, this study resulted identification as can be observed in Table 3. In order to give better illustration, researcher presented the personality type for school programs leading to certain occupations in a spider diagram, such as a skin counselor (SEA). Figure 1 describes the personality type of school programs leading to occupation as skin counselors.

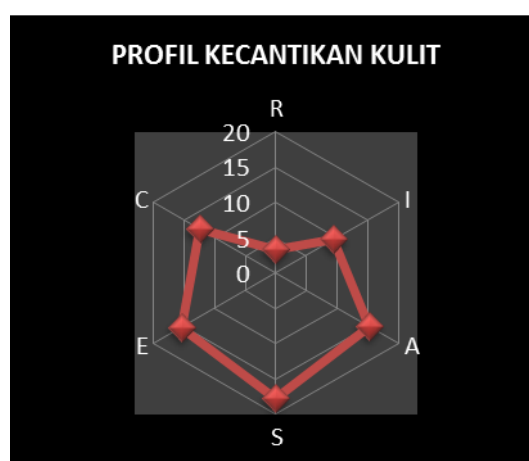


Figure 1. Personality Type for School Programs to be Skin Counselors

From Figure 1, it is easily seen that the three highest scores on the instrument of vocational scales are in the scale Artistic, Enterprising, and Social. This means that if school psychologists equipped with the instrument of vocational scales in their counselling to students of Junior High Schools, they may easily suggest to certain programs in High Vocational Schools. The usage of the instrument is direct and easy to be used by School Psychologists of School Counselors.

At present, the instrument is available in a web system and the personality identified is also directly appeared in the computer monitor used. It is easily printed when such results are needed for any purposes. Any school psychologists or counselors who wish to use the instrument may communicate to the psychometric group at the Faculty of Psychology, The Muhammadiyah University of Surakarta. It is urged to school principals and administration to use the instrument as one of the supplementary consideration in selecting or placing students in a school program at the High Vocational Schools.

IV. CONCLUSION AND RECOMMENDATIONS

Based on the presentation of the results and discussion, researcher concluded that the instrument of vocational scales is provided valid scores for predicting school programs in High Vocational Schools. Therefore, it is recommended to use the instrument as one of indicators in the selection and placement of Junior High Schools graduates to certain school programs with specific track for certain occupations at the High Vocational Schools.

For future development more studies are still needed to validate with other school programs (there are about 86 school programs) in the High Vocational Schools which are still unknown their personality types. Besides that, the workers who are in the occupations are also still

unknown of their personality types. It is still at the beginning of knowing such conditions and systems.

REFERENCES

- [1] J. L. Holland. *Making vocational choices: A theory of vocational personalities and work environments*. Englewood Cliffs, NJ: Prentice-Hall. 1985.
- [2] L. R. Aiken. "Three coefficients for analyzing the reliability and validity of ratings". *Educational and Psychological Measurement*, 1985, 45, pp 131-142.
- [3] Aljufri B. Syarif & Kumaidi. *Minat Kejuruan murid-murid Sekolah Menengah Tingkat Atas di Sumatera Barat. Laporan Penelitian (tidak dipublikasikan)*. Institut Keguruan dan Ilmu Pendidikan Padang, Maret 1990. (High Technical and Vocational School students types of vocational interets. Unpublished Research Reports. The Institutes of Teacher Training and Education of Padang, 1990).
- [4] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. *Standards for educational and psychological testing (third edition)*. Washington, DC: American Psychological Association. 1985.
- [5] C. H. Lawshe. "A quantitative approach to content validity". *Personel Psychology*, 1975, 28, pp 563-575.
- [6] K. G. Joreskog & D. Sorbom. *LISREL 8.80: User's Reference Guide*. Chicago: Scientific Software Insternational. 1996.

RASCH MODEL ANALYSIS FOR PROBLEM SOLVING INSTRUMENT OF MEASUREMENT AND VECTOR SUBJECT

Mustika Wati¹, Yetti Supriyati², Gaguk Margono²

¹Physics Education Department, FKIP Universitas Lambung Mangkurat

²Research and Education Evaluation Department, Universitas Negeri Jakarta
mustika_pfis@unlam.ac.id

Abstract - This research is a development research of instrument to measure students' problem solving skill in measurement and vector subject matter. The purpose of this research is to analyze the problem solving instrument using Rasch model in order to know the instrument and item validity fit to IRT, reliability, and difficulty level. The instrument consists of 10 items in form of essay which is tested to 243 students from SMA Negeri 7 Banjarmasin. The results of this research are: 1) validity of instrument seen from data fit in Item Response Theory (IRT) model is valid with the mean value of INFIT MNSQ 1.00 and deviation standard 0.15 which means that the overall instrument fit the PCM 1 PL model, 2) validity of items is valid with the value of INFIT MNSQ between 0.80 and 1.28 which means that all 10 items are categorized as fit, 3) reliability of problem solving instrument seen from the item reliability estimation is 0.81 (internal consistency) with the mean case estimation of INFIT MNSQ 1.02 and deviation standard 0.71 means that overall testee fit the Rasch model, 4) difficulty level of items are around -1.07 to 0.65 which means that all 10 items are good with the easiest item is number 3 and the most difficult item is number 8.

Keywords: *problem solving, Rasch model, measurement, vector*

I. INTRODUCTION

Learning is a process to have a better change whether in individual characteristics, behavior, skill or knowledge. In all active learning process, the learners learn according to their own needs and pace (Orhan & Ruhan, 2006) [2]. In this research, learners refer to the students who are receiving education or formal learning process. Students are demanded to gain new skill or knowledge to prepare them in facing their future which is full of challenge and new things. While the world keeps changing and problem keeps coming, students have to own a really essential skill to help them facing those problems in the real life which is problem solving skill. Problem solving skill can be measured by instrument which is arranged according to the problem solving skill level of question, but the researcher have to know first whether the instrument is feasible and valid to measure the skill. Therefore, it is necessary to analyse the problem solving skill instrument to make sure that the instrument is able to measure properly in order to reach the research's purposes.

II. MATERIALS

A. Problem Solving

Problem solving method is a method to solve a problem in which generally students are being given a problem that has to be solved by themselves during the learning process. According to the National Science Teachers Association (1985) problem solving is a very essential ability which has to be developed in science learning. Problem solving is an application result of knowledge and procedure towards a problem situation. According to Rebori (2007), "A problem must first be felt, understood, and alternative choices created before effective groups can consider its solution, implement the decision, and evaluate how well the solution is working" [5].

B. Measurement and Vector

Measuring is comparing the measured thing with the base quantity which is defined before. In measurement topic, there are quantity and unit. Quantity is something whose value can be measured and determined, while the unit is needed to understand the certain quantity or measurement of an object. There are 7 base quantities, which are mass, length, time, electric current, temperature, light intensity, and matter quantity. The quantities are divided in two terms, scalar quantity and vector quantity. Vector indicates the magnitude and direction of a quantity. There are three operations of vector: addition, subtraction, and multiplication operation. Every measurement result contains numbers that state the precision of measurement. Those numbers are called as significant figure.

III. RESEARCH METHODOLOGY

The analyzed variable in this research is the instrument which is used to measure the students' problem solving skill in measurement and vector subject seen from the instrument's validity, fit, reliability, difficulty levels.

A. Instrument Analysis

Instrument analysis aims to analyze the validity, fit, reliability, and difficulty levels of the problem solving skill instrument used in this research. Ratumanan (2003) states that validity shows how good the evaluation tool evaluates what should be evaluated [4]. So, the validity depends on how far the precision of evaluation tool in doing its function. The validity of instrument is analyzed from data fit in Item Response Theory (IRT) model. The instrument is stated as fit to the model if after it is calibrated, it is valid and testee is fit to model means that the measurement is valid (Wright & Masters, 1982: 114) [3]. All the aspects including validity, fit, reliability, and difficulty levels are analyzed by using QUEST program for Rasch model analysis.

B. Data Collecting Technic

Data collecting technic is using test and documentation. Instrument try out test was held in three high schools in Banjarmasin which are categorized as high, medium, and low. The selection of the schools is based on the school's ranking in Banjarmasin. The time given for students doing the test was 2 X 45 minutes. The number of testee or respondents is 243 students and the number of questions in the instrument is 10 items consist of essay questions.

IV. RESULT AND DISCUSSION

A. Instrument fit to PCM 1 PL Model

The whole fit test was developed by Adam & Khoo (1996: 30), if the mean of INFIT MNSQ is around 1 and the standard deviation is 0.0, then the whole instrument is fit to PCM 1 PL model [1]. In the research, the mean of INFIT MNSQ is 1.00 and the standard deviation is 0.15, then the instrument is fit to the PCM 1 PL model.

An instrument items are fit to the model or fulfill the goodness of fit if the score of INFIT MNSQ is between 0.77 and 1.30. In the research, the score of INFIT MNSQ is between 0.80 and 1.28 which means that there is no item in the outer side of fit item acceptance limit according to the model, and then all of the 10 items are categorized as fit.

B. Estimation result

The estimation result of 10 items which were done by 243 respondents with the chance level of 0.50 using polychotomus scoring according to PCM 1 PL four categories which is analyzed using QUEST program can be seen on the Tabel 1.

Table 1. Problem Solving Items' Estimation Result on Measurement and Vector Subject

No	Explanation	Items Estimation
1	Mean and Standard Deviation	0.00 ± 0.63
2	Transformed Mean and Standard Deviation	0.00 ± 0.53
3	Reliability	0.81
4	INFIT MNSQ Mean and Standard Deviation	1.00 ± 0.15
5	OUTFIT MNSQ Mean and Standard Deviation	1.01 ± 0.22

C. Reliability

Problem solving instrument reliability is seen from the item reliability estimation which is 0.71 with INFIT MNSQ case mean estimation is 1.02 and standard deviation is 0.71 which means that

overall testee is fit to Rasch model, because this is the result of essay test so then the data is in polycotomus scale. The reliability of the instrument is good with the score of 0.81 with SD adjusted is 0.68, which means that the instrument is relatively stable if it is being used on a broader scale and it is closer to the real respondent characteristics.

D. Difficulty levels

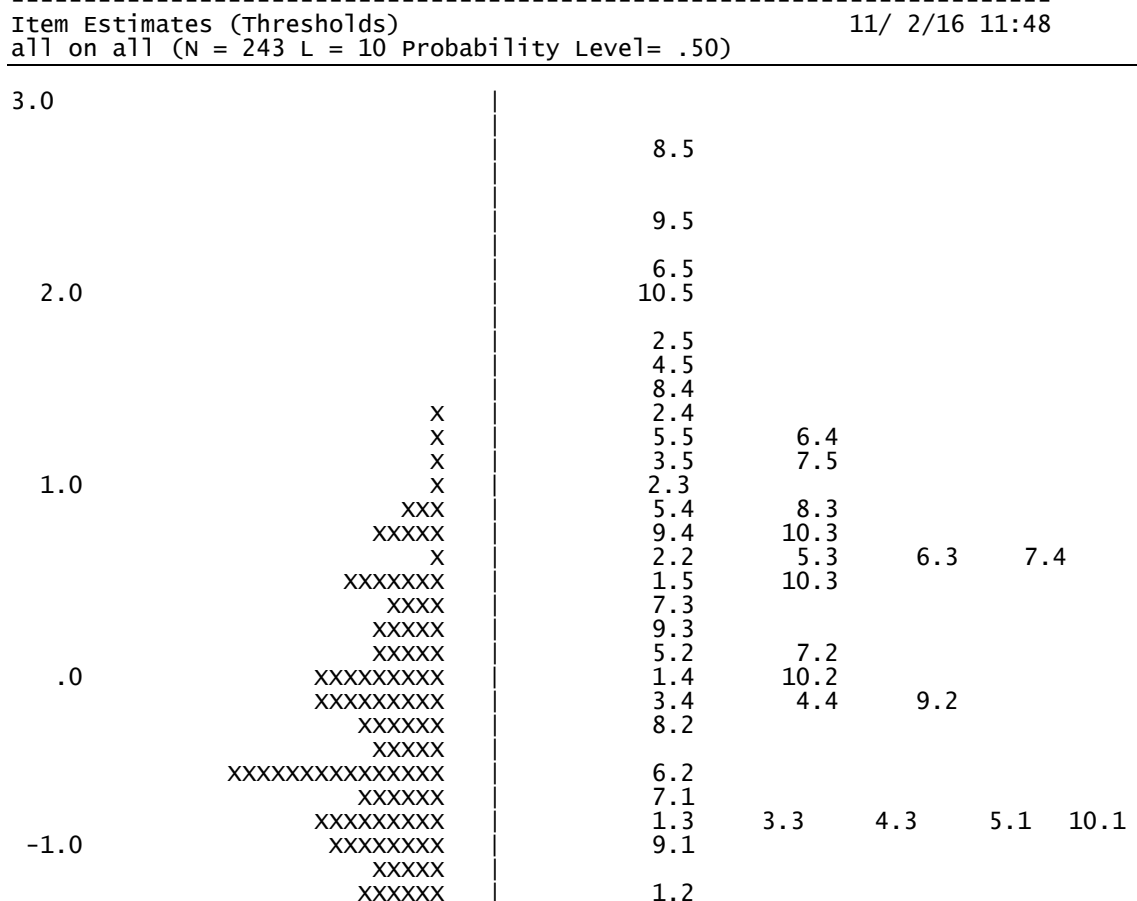
Difficulty levels are analyzed from the graphic which is provided in the figure 1 and figure 2. Items difficulty levels are around -1.07 to 0.65 with mean score of 0 and standard deviation of 0.63. The difficulty levels are between -2.00 and 2.00, so that all of the 10 items are all good category of difficulty levels. The easiest item is item number 3 and the most difficult item is item number 8. Overall the 10 items are able to evaluate the problem solving skill of students while doing the Physics problems. Table 2 shows the list of problem solving skill instrument difficulty levels in order.

Table 2. Problem Solving Skill Instrument Difficulty Levels in Order

Difficulty levels	Item
-1,07	3
-0,80	1
-0,78	4
0,18	6
0,27	7
0,34	5
0,34	2
0,43	9
0,45	10
0,65	8

The following figure shows the item estimation of difficulty levels in which X represents 2 students.

Problem solving



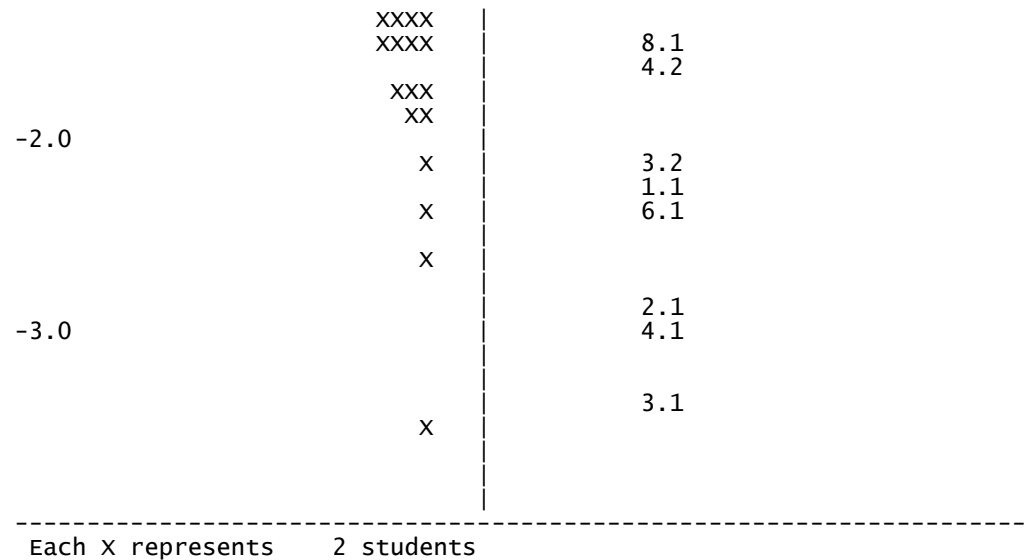


Figure 1. Difficulty Levels of Item Estimation

The figure above shows that the most difficult item is item number 8 and the easiest item is item number 3. While the figure below shows the items acceptance limit.

Problem solving

Item Fit
2/16 11:48
all on all (N = 243 L = 10 Probability Level= .50)

INFIT	.56	.63	.71	.83	1.00	1.20
MNSQ						
1 item 1						*
2 item 2			.		*	.
3 item 3			.		*	.
4 item 4			.		*	.
5 item 5			.	*		.
6 item 6			.	*		.
7 item 7		*	.			.
8 item 8		*	.			.
9 item 9		.	.			*
10 item 10		.	.		*	.

Figure 2. Items Acceptance Limit

Figure above shows that all 10 items are fit to Rasch or 1-PL model with acceptance limit of ≥ 0.77 to ≤ 1.30 .

V. CONCLUSION

The conclusion of the research is that the validity of instrument seen from data fit in Item Response Theory (IRT) model is valid with the mean value of INFIT MNSQ 1.00 and deviation standard 0.15 which means that the overall instrument fit the PCM 1 PL model, the validity of items is valid with the value of INFIT MNSQ between 0.80 and 1.28 which means that all 10 items are categorized as fit, the reliability of problem solving instrument seen from the item reliability estimation is 0.71 with the mean case estimation of INFIT MNSQ 1.02 and deviation standard 0.71 means that overall testee fit the Rasch model, and the difficulty level of items are around -1.07 to 0.65 which means that all 10 items are good with the easiest item is number 3 and the most difficult item is number 8. So, the instrument is feasible to be used to measure the students' problem solving skill in measurement and vector subject in this research.

ACKNOWLEDGMENT

The authors thank the many reviewers who read and provided insight into this paper. Particularly valuable were insights and suggestions from the Research and Educational

Evaluation (PEP) Department, Jakarta State University professors, and also the family members of the authors who always support and help authors in many ways.

REFERENCES

- [1] Adams, R. J. & Khoo, S. T. (1996). *Quest: The Interactive Test Analysis System*. Melbourne: ACER.
- [2] Akinoglu, O. & Ruhan, O. T. (2007). The Effect of Problem- Based active Learning in Science Education on Students' Academic Achievement, Attitude and Concept Learning. *Eurasia Journal of Mathematics, Science & Technology Education*, 3(1), 71-81.
- [3] Benjamin D. Wright & Geoff Masters. 1982. *Rating Scale Analysis: Extends Rasch item analysis and person measurement to attitude questionnaires, performance ratings and any other ordered category data*. ISBN 0-941938-01-8.
- [4] Ratumanan, T. G. and Laurens, T. 2003. *Evaluasi Hasil belajar yang Relevan Dengan Kurikulum Berbasis Kompetensi*. Unesa University Press.
- [5] Rebori, M. K. 2007. Effective Problem-Solving Techniques For Groups. *Journal of Cooperative Extension, University of Nevada*. <http://www.unce.unr.edu/publications/files/cd/other/fs9726.pdf>. Accessed, 28 December 2014.

ANALYSIS OF MATHEMATICAL REASONING ABILITY OF ELEMENTARY SCHOOL STUDENTS USING TIMSS TEST DESIGN

Noening Andrijati

Department of Elementary School Teacher Education, Semarang State university

andrijt06@gmail.com

Abstract - This study aims to investigate the mathematical reasoning abilities of fourth-grade students of Elementary School in Tegal to solve the TIMSS problem. Less information and unavailability of contextual instruments of mathematical reasoning abilities of Indonesian elementary school students in solving TIMSS problems are the reason for doing this study. The approach used in this study is a mixed method, qualitative and quantitative. Data were collected from 106 fourth-grade students of Elementary Schools in Tegal city, which were the partner of Elementary School Teacher Education Department, Semarang State University. Twelve items regarding the reasoning in TIMSS that have been adapted to Indonesian context were tested on the students. The results showed that the reasoning abilities of fourth grade students of Elementary Schools in Tegal were still in the low category. The students had difficulties in solving TIMSS problems in all content domains, especially in data presentation. Meanwhile, there were significant differences in reasoning abilities between students using Curriculum of 2013 and ones using School-based Curriculum.

Keywords: *mathematical reasoning abilities, TIMSS, elementary school students*

I. INTRODUCTION

The low achievement of Indonesian students in some international comparative studies such as the Trends in Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) is one of the considerations of the turn of School-based Curriculum into Curriculum of 2013. The challenges of the future which increasingly complex and global require readiness in the field of education, including curriculum development.

TIMSS is an international comprehensive study of Mathematics and Science achievements or trends of fourth-grade students of elementary school and eighth-grade students of junior high school. Mathematics Achievement of Indonesian students for 5 times Indonesia's participation in TIMSS, namely in 1995, 1999, 2003, 2007, and 2011, have not shown satisfactory results. Indonesian students rank are 35th of 46 countries in 2003, 36th out of 49 countries in 2007, and 38th out of 42 countries in 2011 [1] & [6].

Mathematics TIMSS 2011 items are classified into two domains: content domains and cognitive domains [4]. The content domains are in accordance with Mathematics content in Indonesian Junior High School Curriculum, i.e. number, algebra, geometry shape and measurement, data display and probability. Meanwhile, Mathematics content in Indonesian Elementary School includes the number, geometry and measurement, and data display. Cognitive domains are related to students process of thinking which include knowing facts and procedures, using concept (applying), solving routine problems and reasoning. The reasoning is the expected behavior of students when dealing with mathematical content domains. The percentage distributions of the content domains and cognitive domains of Mathematics for Elementary School are presented in the following tables.

Table 1. Percentage of Elementary School Mathematics Content Domains

Content Domains	Percentage
Number	50%
Geometry and Measurement	35%
Data Display	15%

Table 2. Percentage of Elementary School Mathematics Cognitive Domains

Cognitive Domains	Percentage
Knowing	40%
Applying	40%
Reasoning	20%

The result of TIMSS 2011 showed that the achievement of Indonesian Elementary School students in knowing ability was 37 %, applying ability was 23 %, and reasoning ability was 17%. Being compared with the knowing ability and the applying ability, the reasoning ability is the poorest mathematics ability of Indonesian students. According to [4], mathematical reasoning involves the ability to think logically and systematically. This kind of reasoning includes intuitive reasoning (spontaneity) and inductive reasoning (inference) based on the patterns and regularities that can be used to solve non-routine problems. Non-routine problems are problems that may be very unfamiliar to students; it can be pure mathematics problems or conditioning to everyday life. Domains associated with reasoning involve the ability to analyze, generalize / specialize, integrate / synthesize, justify, and solve non-routine problems [4].

The results of the PISA study in Organization for Economic Co-operation and Development or OECD [6] confirms the TIMSS results, which showed that the level of Indonesian students' mathematical literacy was still relatively low, was under level one. 76.6% of Indonesian students could only use the procedure, formula, and the basic algorithms. This indicated that the abilities of Indonesian students were relatively well in solving problems concerning the facts and procedures, but very weak in solving the non-routine problems relating to justification or giving evidence, solving problems that require reasoning, finding generalizations or conjectures, and finding the relationship between the data or facts discovered.

TIMSS study was conducted using an assessment instrument that applies internationally and did not specifically use the Indonesian context. Several items in the TIMSS used the term, place or event that was not well known by Indonesian students, for example, subway, the name of the city, and type of sport or game.

The context of an assessment affects the assessment result because the context makes the students mindset to remember the concepts that have been learned, to connect them with the existing problems in the context, and then to formulate a solution that is appropriate to the context [3]. Stacey [7] emphasizes the importance of context in learning and assessment as students are prepared for the challenge of the future so they need to be introduced to a variety of contexts that include various aspects of life. Considering the importance of context in the assessment and the absence of studies on mathematics abilities using instruments in the context of Indonesia, it is necessary to study the achievements of the students' mathematics abilities including reasoning ability.

Since 2013, the Indonesian government has applied two curricula, namely the School-based Curriculum (KTSP) and Curriculum 2013. Implementation of these two curricula is different on the emphasis, process standard, and learning assessment. The School-based Curriculum gives more emphasis on the learning aspects of knowledge while the Curriculum 2013 gives more emphasis on balancing of soft skills and hard skills that include aspects of attitudes, knowledge, and skills. The standard learning process in the School-based Curriculum includes exploration, elaboration, and confirmation while the Curriculum 2013 learning process is done using a scientific approach, which includes observing, asking, processing, presenting, concluding, and creating. Because of the emphasis on the learning aspects of knowledge, the School-based Curriculum assessment also emphasizes the aspects of knowledge. The assessment in the Curriculum 2013 uses authentic assessment, which measures all the competencies comprehensively, covering attitudes, knowledge, and skills. Based on these differences, Curriculum 2013 application seems to be giving more opportunities for mathematics learning in

developing reasoning that the students have the ability to think creatively and critically. Nevertheless, the mathematical reasoning abilities of students using the School-based Curriculum and of students using Curriculum 2013 need to be investigated.

Based on this background, the problems to be solved through this research are (a) how far are the students reasoning ability achievements using TIMSS test design?; and (b) are there differences in mathematical reasoning ability of students using different curricula? This study aims at (a) measuring the student mathematical reasoning abilities using TIMSS test design, and (b) obtaining information whether there is any difference between the reasoning abilities of students using the School-based Curriculum and the Curriculum 2013. The research findings will provide important information for scholars, researchers, and policymakers to better understand the in-depth mathematical ability achieved by students for educational improvement.

II. METHOD

This study used a mixed method approach, qualitative and quantitative [2]. The qualitative method was used to describe mathematical reasoning ability while the quantitative analysis was used to test the hypothesis of whether there is any difference in mathematical reasoning abilities based on curricula used.

The sample taken by random sampling was the fourth-grade students of the Elementary Schools which were the partner of Elementary School Teacher Education Department (PGSD) UPP Tegal, Faculty of Education, Semarang State University. The sample of 106 students came from four elementary schools, consisting of 30 students from State Elementary School Mangkukusuman 8 that used Curriculum 2013, and 76 were from State Elementary School Tegalsari 4, State Elementary School Tegalsari 12 and State Elementary School Kalinyamat Kulon 3 that used School-based Curriculum. The sample was fourth-grade students because they had learned the fourth-grade material in the second semester and it was in accordance with the provisions of TIMSS.

Data were collected by taking 12 TIMSS 2011 test items [9] especially the reasoning skills, which had been adapted to Indonesian context, i.e. in terms of language, place, and event. The test consisted of 6 items of number (50%), 4 items of geometry and measurement (33%), and two items of data presentation (17%). Before being tested in the sample students, this test was tried on 60 elementary school students and the test results were analyzed using the Quest program to determine the characteristics of the items and the test. The results of data analysis of this trial showed that the level of item difficulty was good (in the range -1.58 - 1.50), the level of item discrimination was good (in the range 0.17 to 0.64), and the reliability coefficient was high (0.87). Thus, the test has been qualified as a good test. This test was applied to the sample students and then the scores were tabulated scores. The scoring referred to the TIMSS scoring procedure, i.e. a score of 2 for a complete and correct answer, a score of 1 for a correct answer but not complete, and a score of 0 for a wrong answer or no answer.

Data were analyzed qualitatively (descriptive) as well as quantitatively. Descriptive analysis was conducted by converting the scores using the formula:

$$Score = \left(\frac{\text{obtainedscore}}{\text{maximumscore}} \times 700 \right) + 100$$

Then, students value was described by the following categories:

Table 3. Categories of mathematical reasoning abilities based on TIMSS benchmarks

Score	Category
625 – 800	Advance
550 – 624	High
475 – 549	Intermediate
100 – 474	Low

The quantitative analysis was conducted through a comparison test with Independent-Sample T-Test [8], which were used to compare mathematical reasoning skills of students who were learning based on the Curriculum 2013 and on the School-based Curriculum. Data for hypothesis testing was raw scores.

III. RESULTS AND DISCUSSION

In this study, the students' mathematical reasoning skills were demonstrated by students' answers to 12 test items. In the data processing, many students gave answers without explanation and working step in solving such questions. This phenomenon showed that students are less able to provide explanations or arguments on math problems. The achievement of mathematical reasoning skills based on content domain and test results of students reasoning ability differences based on School-based Curriculum and Curriculum 2013 was presented below.

A. Mathematical Reasoning Ability

Data from mathematical reasoning ability tests showed that the students' average score was low (455.78). Specifically, students who received scores of 625 above which included the very high category was only about 9.43%, the high category about 12.26%, the intermediate category about 24.53% and the low category about 53.77%.

Table 4. Percentage of Students Reasoning Ability Based on TIMSS Benchmarks

Score	Category	Percentage of Student
625 – 800	Advance	9.43%
550 – 624	High	12.26%
475 – 549	Intermediate	24.53%
100 – 474	Low	53.77%
Total		100%

Data of the students mathematical reasoning ability were also grouped based on the content domain. The result was showed in the following table.

Table 5. Students Mathematical Reasoning Ability Based on Content Domain

Konten	Students' Average Score	Category
Number	463.2	Low
Geometry Shape & Measurement	471.46	Low
Data Display	402.1	Low

The data in Table 5 showed that the average scores of the three content domains (number, geometry shape and measurement, and data display) were in a low category. This achievement indicated that students were less able to solve the problems in TIMSS mathematical reasoning ability test, although they had been adjusted into Indonesian context. The students had difficulty in solving such problems because their designs were different from those commonly faced; they were non-routine problems.

The smallest mean of the students reasoning ability was data display. Although the problems of data display were quite easy, the students experienced difficulties because they had not learned such problems using the School-based Curriculum. Similarly, the students using Curriculum 2013 only learned the introduction of simple data, fulfilled and read table data. Overall, the achievement of students' reasoning ability based on the content domain is as follows.

Table 6. Comparison of the reasoning ability based on the use of the Curriculum

Content	Mean		Relation
	Curr 13 (M_1)	SBS (M_2)	
Number	605.56	407.02	$M_1 > M_2$
Geometry shape and Measurement	522.92	451.15	$M_1 > M_2$
Data Display	415	397.04	$M_1 > M_2$
Total	546.25	420.07	$M_1 > M_2$

B. Reasoning Ability Difference Based on Curriculum

To find out whether there were differences in mathematical reasoning abilities between the students using Curriculum 2013 and those using School-based Curriculum, Independent-Sample T-Test using SPSS was conducted. The data were the raw scores obtained by the students in solving the 12 test items; the smallest score of 0 and the highest score of 24. The results of the descriptive analysis of the raw score data showed that the average score of mathematics reasoning skills was 15.30 on Curriculum 2013 and 10.97 on School-based Curriculum; the mathematics reasoning abilities average score of the students using Curriculum 2013 was 4.33

points higher than of those using School-based Curriculum. Meanwhile, the standard deviation of Curriculum 2013 (3.35) was smaller than School-based Curriculum (3.67). Furthermore, the results of T-Test can be seen in the following table.

Table 7. The Results T-Test of Reasoning Ability Based on Curriculum Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	Df	Sig. (2-tailed)
Skor	Equal variances assumed	.121	.729	5.610	104	.000
	Equal variances not assumed			5.827	57.712	.000

The results of the analysis in the table showed that the t statistic was 5.610 and the probability of error (p) was 0.00. Because $p < 0.05$, it could be stated that there were significant differences between the mathematical reasoning abilities of the students using the Curriculum 2013 and of those using School-based Curriculum. Based on the average scores, the mathematical reasoning skills of the students using Curriculum 2013 were higher than of those using School-based Curriculum. Therefore, it could be assumed that the mathematical reasoning skills of the students using Curriculum 2013 were better than of those using School-based Curriculum.

The differences of the students' mathematical reasoning skills based on curriculum could be caused by the elementary school learning using Curriculum 2013 which had been implemented well. Attitude, knowledge, and mathematics skills were achieved through meaningful activities covered by the scientific approach (observing, asking, processing, presenting, concluding and creating). Such learning activities allowed students to construct their own knowledge and skills. Learning assessment in Curriculum 2013 was authentic and comprehensive and the instruments used were adjusted to the learning activities undertaken by the scientific approach so that the assessment was appropriate to measure the process and the student performance.

IV. CONCLUSIONS AND RECOMMENDATIONS

Based on the research results and discussion, the research concluded:

(a) The achievement of the ability of elementary school students was still low for all content domains (number, geometry and measurement, and data display), although the TIMSS problems had been adapted to Indonesian context, i.e. in terms of language, place, and event. The students' answers to the test questions were not accompanied by the calculation procedures and or explanations. This condition indicated that students were less able to provide explanations or arguments to the tested mathematics problems;

(b) There were significant enough differences in mathematical reasoning abilities between elementary school students in Tegal using Curriculum 2013 and those using School-based Curriculum.

This research is still limited to a discussion of the achievements of the reasoning abilities of students in elementary schools in Tegal using TIMSS test design. Therefore, further research is still needed to discuss it in more depth, for example, the study of factors that affect students' reasoning ability. In addition, teachers need to develop learning and assessment that students build reasoning, including familiarize students working on the reasoning problems which are internationally designed, e.g. TIMSS.

REFERENCES

- [1] Badan Penelitian dan Pengembangan, Kemendikbud. 2013. *Laporan Studi Kajian Peserta Didik Tingkat Dasar dan Menengah*. Jakarta: Badan Penelitian dan Pengembangan.
- [2] Creswell, J.W. 2003. *Research Design: Qualitative, Quantitative, and Mixed Methods Approach*. Second Edition. California: Sage Publisher. Inc
- [3] Mahdiansyah & Rahmawati. 2014. *Literasi Matematika Siswa Pendidikan Menengah: Analisis Menggunakan Desain Tes Internasional dengan Konteks Indonesia*. Jurnal Pendidikan dan Kebudayaan, Volum 20, No. 4, hlm 452 – 469.
- [4] Mullis, I.V.S.; Martin, M.O.; Ruddock, G.J.; O'Sullivan, C.Y.; and Preuschoff, C. 2009. *TIMSS 2011 Assessment Framework*. Chesnut Hill, MA, USA: TIMSS & PIRLS International Study Center.
- [5] OECD. 2013. *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading, and Science*. (Volum 1). Paris: PISA-OECD Publishing.

-
- [6] Pusat Penilaian Pendidikan Balitbang Kemendikbud. 2012a. *Determinant of Learning Outcomes TIMSS 2011: Final Report*. Jakarta: Pusat Penilaian dan Pendidikan.
 - [7] Stacey, K. 2011. *The PISA View of Mathematical Literacy in Indonesia*. Indo MS J.M.E 2 (2), hlm 95 – 126.
 - [8] Sudjana. 2005. *Metode Statistik*. Edisi Enam. Bandung: Tarsito.
 - [9] TIMSS. 2013. *TIMSS 11 User Guide for The International Database: Released Items Mathematics – Fourth Grades*. Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA): TIMSS & PIRLS International Study Center

THE ACCURACY OF TESTEES' ABILITY ESTIMATION OF THE ESSAY TEST AND TESTLETS IN MATHEMATICS THROUGH THE GRADED RESPONSE MODEL (GRM) APPLICATION

Purwo Susongko, Wikan Budi Utami

Science Education Departement, University of Pancasakti Tegal
Mathematics Education Departement, University of Pancasakti Tegal
(Kusumatirto@gmail.com)

Abstract - The purpose of this study are to find: (1) The accuracy of estimation of the testees' ability of mathematics achievement of the essay test and testlets through the Graded Response Model (GRM) Application, (2) The precision of estimation of the testees' ability of mathematics achievement of the essay test and testlets through the Graded Response Model (GRM) Application. This study was conducted in two stages, empirical and simulation. Empirical studies conducted as the mathematics achievement test of 277 Year X students SMA III Slawi on 2015. The testees' ability parameter of the estimation results of empirical study are used to generate true parameter of testees' ability. The data were generated from the parameters of the estimation results of the empirical study. The data were generated on the basis of the sample sizes of 500 testees and the numbers of items were 10. Replication is done 25 times, using WinGen 3.0 program. Item Parameters generated on the standard normal distribution. The parameters of testees' ability estimated by using a computer program Multilog Version 7:03. Precision in parameter estimation used to mean squared error (MSE). To assess the accuracy the coefficient of Pearson Correlation was calculated. The findings of the study are: (1) The accuracy of the estimation of testees' ability of mathematics achievement essay test is higher than testlets, (2) The precision of the estimation of testees' ability of mathematics achievement essay test is higher than testlets.

Keywords: *accuracy, precision, testees'ability, essay test , testlets*

I. INTRODUCTION

Test can use the collection of multiple items in a group are interconnected. A set of interrelated items which are part of a test commonly called testlets (Wainer, 1995). Testlet been used in the measurement in the field of language (Lee et al, 2001), TOEFL (Wainer & Wang, 2000), law school admission test (Wainer , 1991), medical school admission test (Zenisky et al, 2002), the measurement on a large scale (Ferrara et al, 1997) and attitudes toward racial identity (Fischer & Tokar, 1998). In the TOEFL test, testlets mainly used in the reading. Testlets also widely used in the world of public health and medical (Hamilton, Clayon B, et al 2015).

Reported by Zenisky, Hambleton, & Sireci (2002) that the item in one testlets not assume the character of local independence so that when these items are considered as items that are independent then it does not meet the requirements analysis to item response theory (IRT). Lee, et al. (2001) did equalize using polytomous item response theory, on a test composed of testlets. Lee, et al. (2001) showed that better psychometric equivalency using collection of items as a single response that is polytomous than items is considered as the independent items.

Many studies showed that the response of testlets to testlets more good in psychometric view when modeled in polytomous response that are multikategorik. Some experts have devised

theoretical support to the application of polytomous IRT on testlets scoring (Verhelst & Verstralen, 1997). Wainer, Bradlow & Wang (2007) developed a theoretical framework for the analysis of items test based testlet then called Testlets Response Theory (TRT). The theoretical framework shows that testlets can be analyzed using polytomous IRT. It provide the possibility the test items in an essay test with that in testlets that have many advantages .

There are many benefits when tests are administered in testlets. In addition to the advantages of objective tests in general, testlet also has a scoring system that is multikategorik. Their similarity of IRT analyzes in essay test and testlets raises the question of the extent to which differences in the effectiveness of the testlet and essay test. Some studies do a comparison of both of the test, but is limited to non psychometric aspects of the classical theory approach so as not to provide information about the measurement accuracy of both of the test. By comparing the effectiveness in psychometric view will be known in accuracy when used as a instrument.

Purwo Susongko (2009) conducted a comparison of the effectiveness of the essay test and testlets and through graded response model application. The results showed that the empirical and simulation, the average value of the function information item on the essay test higher than the average value of the function item information in the testlets. This implies that the essay test is more effective than testlets from the informationzitem function is generated.

The study does not answer about the accuracy of the both test to estimate on the testees' ability. The accuracy of testees' ability estimation is parameter of the effectiveness in psychometric view because basically the test is to determine the testees' ability. Further, it needs research about differences in the accuracy of estimates of the testees' ability in an essay test and that in testlets. Based on the background, the problem in this research are as follows:

1. Is there a difference in the accuracy of estimation the testees' ability on mathematics achievement test in an essay test with that in testlets through the Graded Response Model (GRM) Application?
2. Is there a difference in the precision of estimation the testees' ability on mathematics achievement test in an essay test with that in testlets through the Graded Response Model (GRM) Application?

II. RESEARCH METHOD

The development of the instrument initiated by compiling mathematics achievement tests as mathematics achievement test for Year X students of Senior High School. The instrument consists of 5 essay items. With the same test items, the essay tes was converted into 5 testlets. The number of items was adjusted to the material coverage and the time allocation which was 60 minutes. Mathematics achievement test was made by researchers in collaboration with the Mathematics teachers in SMA III Slawi. Both of the tests is admisnisterated on 277 students of class X in SMA III Slawi.

To obtain evidence of the content validity or content representation, the test papers were assessed through the expert judgment. To prove the construct validity and the unidimensional assumption of the tests developed, the data obtained from the tryout results were analyzed by using the structural equation modeling (SEM). The SEM employed to prove the construct validity and the unidimensional assumption was the confirmatory factor analysis (CFA) model. The program employed to analyze the data using this model was PRELIS to obtain the product moment correlation coefficients among items. The scoring of the two instruments employed the analytic method involving four criteria, namely scores 0, 1, 2, and 3. The answer to the previous stages affect the next stage, so that students who answered correctly on the first stage was given a score of 1 students who can answer correctly the whole stage, was given a score of 3 students who answer correctly the second stage, but the first one stage or in two stages of students answered incorrectly given a score of 0. scoring guidelines (rubric answer) created by step completion of each item and have been discussed with team (researchers and teachers). First rater and second rater respectively mathematics teacher who taught in SMA III Slawi Year X.

Testees' ability are estimated using a computer program Multilog Version 7:03. The program is selected because it is easy to use, easy to understand the reading results of the analysis. To determine the accuracy of the testees' ability parameter estimation of essay test and testlet conducted simulation studies. Empirical studies used to determine the value of the testees' ability parameter. Through simulation studies, the testees' ability parameter of empirical study are used

to generate the true score of testees' ability for each test. The data were generated on the basis of the sample sizes 500 testees and the numbers of items were 10. Replication is done 25 times, using WinGen 3.0 program. Item parameters were generated from the item through simulation using the standard normal distribution.

Simulation data used to estimate of the testees' ability. To determine the precision of the estimates used the mean squared difference between the true value and the estimates across replication, referred to as the Mean Squared Error (MSE). As for knowing the accuracy of the estimation used Pearson's r correlation between the true score and the estimates across replication.

With smaller of MSE average value of testees' ability estimation showing that the test form is precise, while larger values of MSE average value of testees' ability estimation showing that the test form is not precise. From correlation coefficient between true score and the testees' ability simulation results can indicate the extent to which the accuracy. Greater the mean of correlation coefficient between true score with the testees' ability simulation results show that more accurate estimates of testees' ability

III. RESEARCH FINDING AND DISCUSSIONS

The empirical data were collected through students' responses to the essay test and the testlets administered to 277 Year I students of the mathematics and science program (MIA) and Social studies Program (IPS) in SMA Negeri 3 Slawi. The test was carried out on May 25, 2015 at 8:00 to 10:00 pm. Construct validity test has been conducted on both the test by using Structural Equation Model (SEM) analysis. SEM analysis is used for the confirmatory validity analysis. Path diagram for the construct validity analysis results of the testlets can be seen in Figure 1, while of the essay test can be seen in Figure 2.

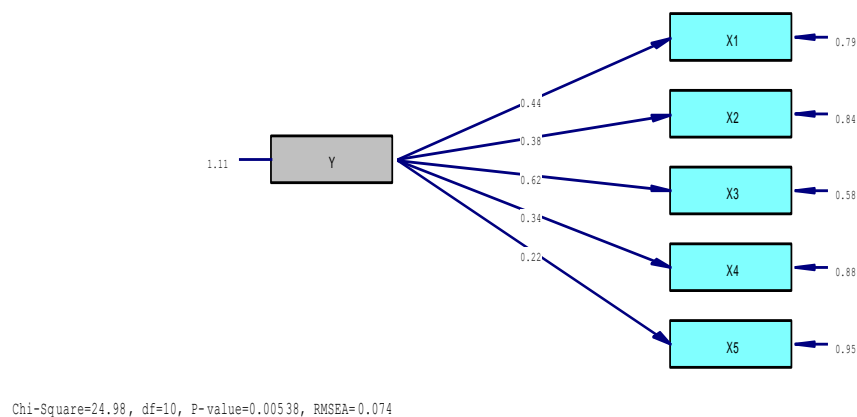


Figure 1. Path Diagram of construct validity analysis on testlets

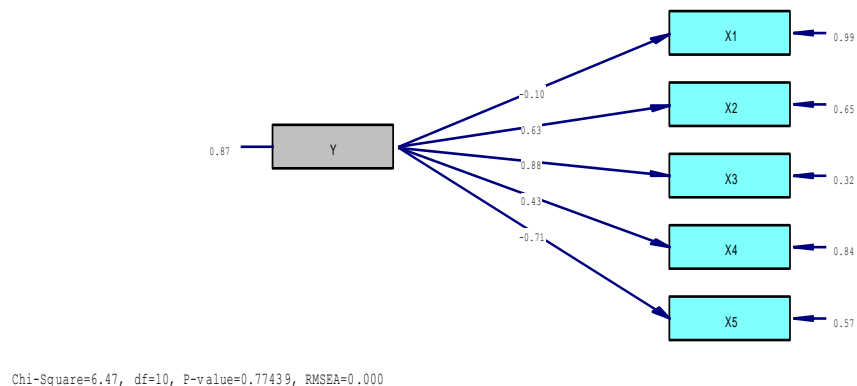


Figure 2. Path Diagram of construct validity analysis on essay test

Of the two pictures can be seen that the model used can be considered valid at least with the following considerations: (1) the value of the ratio chi-square of the degree of freedom of each is

2.498 and 0.647, the value of both are much lower than the cutoff values suggested by Wheaton (1977) that is equal to 5., (2) the value of RMSEA of essay test is 0,000 that is far from the maximum values suggested by Joreskog & Sorbom (1996) that is equal to 0.05, while in the testlets is 0, 074, that the value not too extreme and is still considered to be rational when compared with those criteria.

The following is the item parameters of the estimation results of the empirical study of the testlet (Table 1) and the essay test (Table 2).

Table 1. Item Parameters of the Testlet by Empiric

Item Number	b 1	b 2	b 3	a
1	-1.72	1.07	1.32	0.73
2	-1.10	0.00	1.10	1.10
3	-9.46	-6.19	1.77	0.31
4	-1.10	0.00	1.10	1.00
5	0.35	0.59	0.96	1.08

Table 2. Item Parameters of the Essay Test by Empiric

Item Number	b 1	b 2	b 3	a
1	3.37	-3.60	3.13	5.24
2	-1.10	0.00	1.10	1.00
3	-0.11	3.85	26.27	0.96
4	-1.10	0.00	1.10	1.00
5	3.18	-4.13	4.10	6.46

The estimation of testees' ability in the empirical study in the testlet in the interval between 1.33 and 1:33. While the estimation of testees' ability in the empirical study in the essay test in the interval between -1.33 and 0.14. The estimation results of testees' ability in the empirical study is used to generate the true parameter with the sample sizes of 500.

The true from the testlets and the essay test is used to generate the simulation data with test length by 5 items and sample sizes to 500. Each set was *replicated* 25. Correlation between the testees' ability in simulation and true score for each replication in the essay test and testlets are shown in Table 3. RMSEA value for the testees' ability in the testlet and essay test are shown in Table 4.

As shown in Table 1 and 2, the estimation results of testees' ability in the empirical study show that the use of testlets are overestimate than essay test. It is seen that the testees' ability estimation on the testlet ranged from -1.33 to 1.33 while on the essay test, the testees' ability estimation ranged from -1.33 to 0.14. The distribution of testees' ability in the essay test is 40% in -1.33. This means empirically, testees' ability appear higher on the testlets than on the essay test thus testlets is considered easier than essay test.

Table 3. Correlation between the testees' ability in simulation and true score for each replication in essay test and testlets

Number	replication	Testlets	Essay test
1	1	0.420529	0.656814
2	2	0.465972	0.617115
3	3	0.408278	0.593128
4	4	0.452014	0.661796
5	5	0.472515	0.643351
6	6	0.442672	0.642035
7	7	0.41369	0.616228
8	8	0.38573	0.590539
9	9	0.471735	0.644028
10	10	0.320372	0.642261
11	11	0.416838	0.600648
12	12	0.431693	0.627185
13	13	0.427089	0.648804
14	14	0.376792	0.627232
15	15	0.400892	0.656649
16	16	0.442837	0.633983
17	17	0.451999	0.657832
18	18	0.428893	0.638573
19	19	0.455322	0.631792
20	20	0.406328	0.604489
21	21	0.392176	0.649054
22	22	0.428585	0.627664
23	23	0.397294	0.634545
24	24	0.45824	0.640807
25	25	0.42715	0.650658
	mean	0.423825	0.633488

Table 4. Value of RMSEA of Testees' Ability simulation with the testlet and the essay test

Number	replication	testlets	essay test
1	1	0.499153	0.153767
2	2	0.47228	0.16573
3	3	0.496778	0.160405
4	4	0.479911	0.154997
5	5	0.464823	0.158881
6	6	0.4824	0.152047
7	7	0.491672	0.166123
8	8	0.510259	0.16715
9	9	0.471907	0.160146
10	10	0.539345	0.154061
11	11	0.500709	0.165557
12	12	0.491175	0.161533
13	13	0.490226	0.162092
14	14	0.511503	0.158744
15	15	0.503686	0.156523
16	16	0.487045	0.158521
17	17	0.481408	0.151148
18	18	0.49154	0.15751
19	19	0.477035	0.155428
20	20	0.500337	0.164577
21	21	0.505569	0.158753
22	22	0.487836	0.162303
23	23	0.509775	0.154177
24	24	0.476072	0.152913
25	25	0.488296	0.156561
	mean	0.49243	0.158786

As shown in Table 1 and 2 that the item parameter estimates to the essay test is much more difficult than the testlets. It can conclude that the empirical estimation of item parameter with essay test is irrational until one reaches 26,27. However, these results still credible because the test is made in criteria reference test so that item parameter can have extreme values.

Distribution of true score is still relatively similar to the results of empirical parameter that is to testlets from about -1.33 to 1.33 and in essay test from about -1.33 to 0.14. The mean of true score on testlet is 0.0289 with a standard deviation is 0.03, while the mean of true score in the essay test is -0.587 with a standard deviation is 0.01. Table 3 shows that the correlation coefficient between true scores and parameters of the simulation studies, the essay test tends to have values which are higher than testlets. The mean of correlation coefficient of true score and simulation parameter on the essay test is 0.633 while the testlets is 0.423. This shows that in terms of accuracy, more precise in an essay test than testlet on estimating the testees' ability.

Table 4 shows that the MSE index of 25 replication in simulation studies, the essay test tends to have lower than testlet. The mean of MSE on the essay test of 0.15 while in the testlets is 0.49. This shows that in terms of precision, more precise in an essay test than testlet on estimating the testees' ability.

From the analysis it can be concluded that the essay test more precise on estimating the ability of testees' ability. The results are consistent with several previous studies. Purwo Susongko (2010) proved that essay test is more effective for chemical achievement test than testlets with the application of Graded Response Model. The study was conducted empirical and simulation using the item information functions as a criterion of effectiveness.

Through his studies, Zidner (1987: 607) conclude that the essay test requires a high ability to organize a response, requires the ability to recall the material, requires integrative knowledge and the ability to write well. In the multiple choice test is not found anything like that, because the testee just choose the option that has been prepared. If a testee answered correctly for the same items on the multiple-choice test, it is difficult to presume that the option based on the results of complex thinking.

Kuechler and Simkin (2003: 394) through their study concluded that in the multiple choice test, students have a chance to guess the correct answer is greater than in the essay test. Shepard, (2008: 604), through a study conducted by the National Mathematics Advisory Panel of more than 15 studies, concluded that: (1) an error when the multiple choice test and essay test used to measure the competence of the same, (2) the essay test is used to measure the higher ability of students and (3) the essay test has more information than multiple choice test.

IV. CONCLUSIONS AND RECOMMENDATION

A. Conclusions

1. The accuracy of the estimation of testees' ability of mathematics achievement essay test is higher than testlets,
2. The precision of the estimation of testees' ability of mathematics achievement essay test is higher than testlets

B. Suggestion

1. It needs further research about differences in the accuracy of the essay test and testlets in estimating the testees' ability by considering other variables such as the length of the test, the sample size and the number of replication.
2. Need more in-depth research about differences in the accuracy of the essay test and testlets in estimating the testees' ability with item response theory modeling of other types such as the modeling GPCM, PCM and other models politomos models.
3. Keep a study similar to the field of educational measurement other areas such as in the field of language studies, science and social sciences.
4. In practice, a essay test should be more widely used than the form testlet in educational measurement.

REFERENCES

- [1] Ferrara, S, Huynh, H, & Baghi, H. (1997). Contextual characteristics of locally dependent open-ended item cluster in a large-scale performance assessment. *Applied Measurement in Education*, 10(2),123-144
- [2] Fischer, A. R & Tokar, D. M. (1998). Validity and construct contamination of the racial identity attitude scale-long form. *Journal of Counseling Psychology*, 45(2),212-224.

-
- [3] Hamilton, Clayon B; Maly, Monica R; et al (2015), Health and Quality of Life Outcomes 13 Validation of the Questionnaire to Identify Knee Symptoms (QulKS) using Rasch analysis <http://search.proquest.com/docview/1780120342?accountid=62691>
 - [4] Heri Retnowati (2006). Stabilitas estimasi Parameter pada Regresi Logistik (suatu penerapan pada pengukuran). <http://eprints.uny.ac.id/7246/1/PM-9%20-%20Heri%20Retnowati.pdf>
 - [5] Lee, G.et al. (2001). Comparisons of dichotomous and polythomous item respons models in equiting scores from test composed of teslets. *Applied Psychological Measurement* , 25 (4), 357-372.
 - [6] Purwo Susongko, (2009). Studi Komparasi Kualitas Butir Tes Prestasi Belajar Matematika Pada Tes Bentuk Uraian Dan Bentuk Testlet .Penelitian Mandiri. tidak diterbitkan
 - [7] Purwo Susongko, (2009).Perbandingan Keefektifan Bentuk Tes Uraian dan Testlet Dengan Penerapan Graded Response Model (GRM), Disertasi, Tidak diterbitkan, Pascasarjana UNY Yogyakarta, 2009
 - [8] Purwo Susongko, (2014). Perbedaan ketepatan estimasi tingkat kesukaran butir tes pilihan ganda pada penskoran koreksi dan penskoran konvensional dengan penerapan model Rasch. Makalah tidak diterbitkan. Dipresentasikan pada Seminar nasional Matematika dan pendidikan Matematika, USD Yogyakarta, 13 September 2014.
 - [9] Risky Setiawan (2014) Analisis Dan Simulasi Dengan Program Win-Gen(Strategi Dalam Mengkonstruk Instrumen Soal). e-journal.ikip-veteran.ac.id/index.php/pawiyatan/.../8
 - [10] Shepard, L.A.(2008). Commentary on the national mathematics advisory panel recommendations on assessment. *Educational Researcher*, 37(9),602-609.
 - [11] Van der Linden, W. J & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York ,NY:Spring-Verlag, Inc.
 - [12] Verhelst, N. D & Verstralen, H. H. F. M. (1997). Modeling sums of binary responses by the partial credit model. *Measurement and Research DepartementReports*, 1-18
 - [13] Wainer, H. (1995). Precision and differential item functioning an a teslets based test : The 1991 law school admissions test as an example. *Applied Measurement in Education*, 8, 157-186.
 - [14] Wainer, H, Sireci, S. G & Thissen, D. (1991). Differential testlets functioning: Definition and detecting . *Journal of Education Measurement*, 24, 185-201.
 - [15] Wainer, H & Wang, X. (2000). Using a new statistical model for teslets to score TOEFL.*Journal of Educational Measurement*,. 37, 203-220
 - [16] Zedner, M.(1987). Essay versus multiple-choice type classroom exams: the student perspective. *Journal of Educational Research*, 80(6),352-358.
 - [17] Zenisky, A., Hambleton, R & Sireci, S. G (2002). Identification an evaluation of local item dependencies in the medical college admission test, *Journal of Educational Measurement*, 39 (4), 291-309

THE COMPARISON OF LOGISTICS MODEL ON ITEM RESPONSE THEORY: 1 PARAMETER (1PL), 2 PARAMETERS (2PL), AND 3 PARAMETERS (3PL)

Rida Sarwiningsih¹⁾, Heri Retnawati²⁾

Department of Educational Research and Evaluation, Postgraduate Program UNY
rida.sw59@gmail.com¹⁾, retnawati_heriuny@yahoo.co.id²⁾

Abstract - This research aims to determine the test items characteristics of National Standard Examination in Chemistry, with a logistics model of 1 parameter (1PL), 2 parameters (2PL), and 3 parameters (3PL) on item response theory. Item analysis was conducted to show the power of an item in choosing the most fit model of a test and comparing the items characteristics use the logistics model of 1 parameter (1PL), 2 parameters (2PL), and 3 parameters (3PL). This research was conducted by using a exploratory descriptive method. Data source based on student's response on High Stakes Testing in Chemistry in Jambi Province at the academic year of 2014/2015. Analysis of data using the item response theory approach with the 1 PL, 2 PL, and 3 PL models assisted by the BILOG Program. The results showed that the number of items that can be said to be a good item in 1 PL model amounted to 39 and there is only one item that is not good on this model. Then, in the 2 PL model, there are 30 items characterized as a good item and 10 items are not good. While in the 3 PL model, there are 38 good items and 2 items were not good. So, in this test, the most fit model analysis is 2 parameter logistics (2PL).

Keywords: *item response theory, logistics model of 1 PL, logistics model of 2 PL, and logistics model of 3 PL*

I. INTRODUCTION

A. CONTAINING THE BACKGROUND

Evaluation is an important stage in the learning activity. In the learning activity is always done for all aspects of learning evaluation in education. One of the evaluation conducted in cognitive aspects that measure used instrument test. Several methods developed for analysing item tests include classical test theory approach (CTT) and item response theory (IRT).

Item response theory (IRT) was developed to overcome the shortcomings of classical test theory (CTT). Hambleton & Swaminathan (1985) and Hambleton, Swaminathan, & Rogers (1991) stated that the main assumptions of this theory are: a) local independence, probability of correct answer one with other items are independent, b) unidimension, substance being measured is unidimension, c) Invariansi parameter means items characteristic does not depend on the distribution of the parameters of the students ability and the parameters that characterize the student is not dependent of the characteristic item.

Previous research emphasizes analysis of item characteristics only based on classical test theory or just based on item response theory. Therefore, this research focuses on discussing about characteristics of items based on item response theory (IRT). There are 3 logistics model in item response theory. They are one parameter (1PL), two parameters (2PL), and the three parameters (3PL). In this research, logistics model in item response theory is compared to known item characteristics and determined the fit model used on the high stake testing chemical in Jambi province school year 2014/2015. Some of the information obtained on the third logistics model consists of several different resources such as item parameters (a) discrimination power, (b) difficulty index, and pseudo guessing (c).

Difficulty index of item parameters is one of item parameter that is very useful in analyzing a test. Allen and Yen (1979) stated that in general the difficulty index of item in the interval -2.0 2.0. On this interval, information about students ability will be obtained for the maximum. In designing the difficulty index of test must consider the purpose of the preparation of the test. A high index difficulty indicates that an item more difficult nor vice versa.

Discrimination power is a matter to distinguishing item ability to difference ability students who haven't and have understood the lesson based on item problem undertaken by students. Based on discrimination power, item can be known whether the item reserved good, revised or rejected. If a item does not distinguish the two students ability, there are possibility that the answer keys are less precise, item has 2 or more correct answers, the competence of which measured less obviously, pseudo guessing does not work, the lesson is very difficult so that students ask a lot of the guessing out, and most students understand about lesson thinking there is missinformation on the item test. At the item test, discrimination power is said to be good if it is greater than or equal to 0,0 until 2.0 (Humbleton & Swaminathan, 1985). Criterion of discrimination power items not too need to be a concern, as long as not negative (Ebel & Frisbie, 1986; Frisbie, 2005). If the value is small, indicating that the slope of score distribution from the population, which also resulted in the validity of the test be low. The value of discrimination power is high shows item have an increasingly good discrimination power while the value of discrimination power is low shows item has not good discrimination power.

Effectiveness of distractor determined to shows the distractor have functioning properly. Effectiveness of distractor can determine based on two criteria, they are distractor chose by low ability students and professional distractor (Saifuddin Azwar, 1999). According Djemari Mardapi (2008) that every chosen answer must chose by at least 5 % of students.

Item response theory have 3 logistics model that used to analyze items, they are logistics model 1 parameter (1PL), logistics model 2 parameter (2PL), and logistics model 3 parameter (3PL). Hou et al (2014) opinion is the difference of the third logistics model on parameters that used to describe items characteristic with the model that used to analyze items test. 1 parameter logistic model (1PL) are often called Rasch model, used to analyze the data that just focuses on one of the item characteristics parameters namely index difficulty, while other parameters such as discrimination power being equal (Hambleton, Swaminathan, & Rogers, 1991). Model logistics 2 parameters (2PL) is used to analyze the data using two parameters namely index difficulty and discrimination power. The item guessing is considered equal to zero. 3 logistic model (3) is parameters used to analyze the data using the parameters of index difficulty, discrimination power, and pseudoguessing. Pseudoguessing is low ability students to correct answer of item with different index difficulty. Value range pseudoguessing is >0.25 . A high value for pseudoguessing indicate the more likely guessed answers correct, whereas the lower value of pseudoguessing, then the less likely to guess the correct answers.

The problem that have shown became a reference for analyzing items on high stakes testing chemistry 2014/2015 in Jambi province. Items are analyzed to demonstrate the characteristics test to compare logistics model 1PL, 2PL, and 3PL on item response theory and it can be determined the most fit model on on high stake testing chemistry 2014/2015 in Jambi province.

B. RESEARCH PROBLEM

Research problem in this research are:

1. How item characteristics on high stakes testing chemistry in Jambi Province based on logistics model in item response theory?
2. What fit logistics model that applied to analyze items in hig stakes testing chemistry 2014/2015 in Jambi province?

C. OBJECTIVES

This research aims to determine the test items characteristics of National Standard Examination in Chemistry, with a logistics model of 1 parameter (1PL), 2 parameters (2PL), and 3 parameters (3PL) on item response theory. Item analysis was conducted to show the power of an item in choosing the most fit model of a test and comparing the items characteristics use the logistics model of 1 parameter (1PL), 2 parameters (2PL), and 3 parameters (3PL).

D. SIGNIFICANCE OF RESEARCH

This research is used to compare items characteristics with logistics model 1PL, 2PL, and 3PL on item response theory. Logistics model on item response theory compared to determined the fit model to analysis the test.

II. RESEARCH METHOD

This research was conducted by using a exploratory descriptive method. It is describe and analyze response students in chemistry on high stakes testing Jambi province. This research analysing results of high stake testing in chemistry based on answers response to knows item characteristics. Items charakteristich analyzed by logistics model 1PL, 2PL, and 3PL to determine difficulty index, discrimination power, and pseudoguessing and then to describe chi-square value on specific degrees of freedom to determine model fit that used on the test.

A. CONTAINING THE APPROACH USED

This research use quantitative approach to analyze students response based on item response theory. This quantitative approach use to determine the parameters that used on the item response theory logistics model.

B. POPULATION AND SAMPLE

Population in this research is participants of high stake testing senior high school sains class in Jambi province as much as 27,100 students. The sampling method used in this research was Simple Random Sampling, namely the determination of the research sample done at random. This method is an equal opportunity to be selected for each individual unit or in the overall population. The number of high schools in the province of Jambi is 162 school. The sample was selected based on a package of UN testing in chemical consists of five packages. Packages reserved used is UAE1110 packages with the number of participants working on that package that is as much as 710 students.

C. SOURCE OF DATA

Source of data can get from PUSPENDIK. It is item response data high stake testing participants in chemistry senior high school 2014/2015 on Jambi province.

D. TECHNIQUE OF DATA ANALYSIS

Technique of data analysis is use item response theory that have 3 logistics model. They are logistics model 1PL, 2PL, and 3PL.

1PL model is:

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}}; \text{ with } i = 1,2,3, \dots, n \quad (1)$$

$P_i(\theta)$: probability test participants who have ability θ to randomly selected item i can answer correctly

θ : level of ability subject tingkat (as independent variable)

b_i : i item difficulty index

e : natural numbers whoses value approaches 2,718

n : the number of items in the test

2PL model is:

$$P_i(\theta) = \frac{e^{a_i(\theta-b_i)}}{1+e^{a_i(\theta-b_i)}}; i = 1,2,3, \dots, n \quad (2)$$

a_i = item discrimination power

3PL model is:

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{a_i(\theta-b_i)}}{1+e^{a_i(\theta-b_i)}}; i = 1,2,3, \dots, n \quad (3)$$

c_i = pseudoguessing

Analysis of data using the item response theory approach with the 1 PL, 2 PL, and 3 PL models assisted by the BILOG MG Program. Item analysis with BILOG MG Program give information about item parameters on classically is phase 1 output. Phase 2 output to get information about discrimination power, difficulty index is analysis in item response theory, and phase 3 output to determine estimation of test participants ability. If there are several item that give negative high biserial correlation, items must be removed so as not to interfere with the

subsequent analysis. Items analysis on item response theory have two parameters, it is item parameter and trait parameter. The trait parameter θ expressed characteristic trait of the participants with the ability of the θ parameter, while the item parameter is revealed through a fit logistic model.

III. RESEARCH FINDINGS AND DISCUSSIONS

This research items test high stakes testing in chemistry 2014/2015 in Jambi province analyzed by item response theory approach with logistics model 1PL, 2PL, and 3PL. Items are multiple choice with 40 item test. Based on documentation data obtained from PUSPENDIK as much as 710 respondents. Analysis of modern item or better known as item response theory. Items were analyzed with statistics program, it is BILOG MG with logistics model 1PL, 2PL, and 3PL. Analysis using BILOG MG results 3 phase. Items characteristics are analyzed in phase 2 and ability student in phase 3. After analysis of items characteristics, then compare 3 logistics model in item response theory to determine the fit model that can use in high stakes testing for chemistry senior high school 2014/2015 in Jambi province. Following the results of analysis item characteristics based on model logistics 1PL, 2PL, and 3PL.

A. Item Analysis model logistics 1PL

The logistic model 1PL involves one parameter that is the difficulty index of item with discrimination power being equal. Items analysis results shown in the following table.

Table 1. Difficulty Index of Item

Difficulty Index		
Category	Total	Percentage (%)
Good	21	52,5
Not Good	19	47,5
Total	40	100

Criteria specified difficulty index between -2,0 until 2,0. Based on the table note that results of analysis using model of logistics 1PL, there are 21 good item or about 52.5% from all items that has examined based on the item difficulty index. While the rest categorized items is not good as much as 7 item or about 47,5%. Item categorized is not good because it has a index difficulty less than -2,0 or more than 2,0. On the 1PL model 1PL index difficulty on chemistry test highest amounted to 1.941 on items 22. Item characteristicss curve for item 26 shown in the following curve.

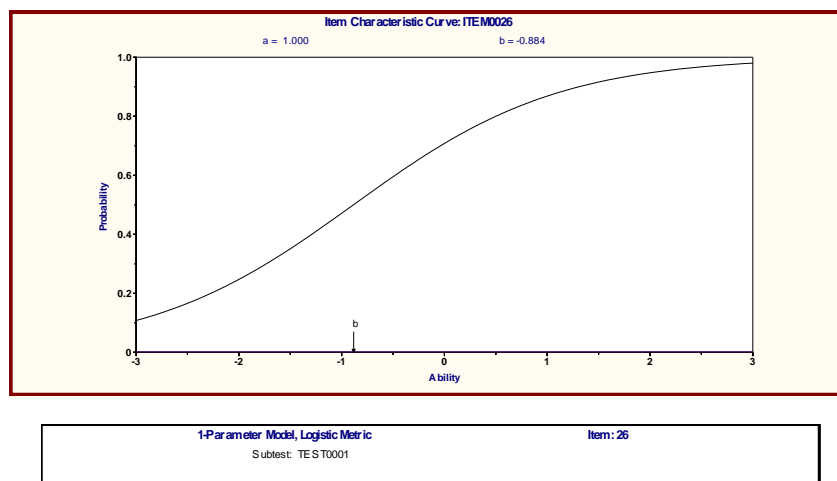


Figure 1. Curve shown an ICC 1PL model logistics on item 26 with $a = 1$, and $b = -0,884$.

B. Item Analysis model logistics 2PL

The logistic model 2PL involves two parameter that is the difficulty index of item and discrimination power with pseudoguessing being equal. Items analysis results shown in the following table.

Table 2. Difficulty Index of Item

Difficulty Index		
Category	Total	Percentage (%)
Good	28	70
Not Good	12	30
Total	40	100

Based on the table note that results of analysis using model of logistics 2PL, there are 28 good item or about 70% from all items that has examined based on the item difficulty index. While the rest categorized items is not good as much as 12 item or about 30%. On the 2PL model index difficulty on chemistry test highest amounted to 1,25 on items 6 and 18. It is assumed that item is difficult item.

Table 3. Discrimination Power of Item

Discrimination Power		
Category	Total	Percentage (%)
Good	35	87,5
Not Good	5	12,5
Total	40	100

Based on table known analysis results with 2PL model logistics there are good discrimination power of item as much as 35 item or 87,5% from all items that has examined, and then not good discrimination power of item as much as 5 item or 12,5%. Discrimination power categorized is not good because it has negative discrimination power value or less than 0,0 or more than 2,0. In this model average the discrimination power value about 0,512 and more highest about 1,813. That discrimination power value is already included in the good categories discrimination power value in accordance with predetermined criteria. Item characteristics curve for item 10 shown in the following curve.

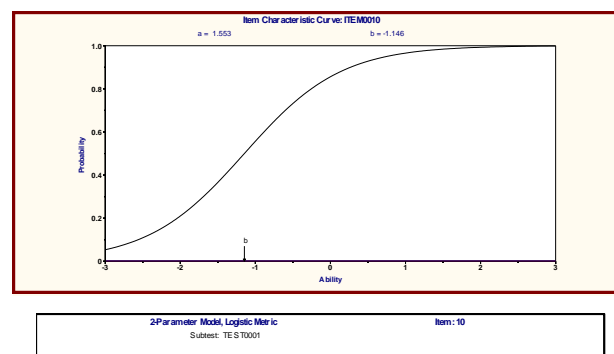


Figure 2. Curve shown an ICC 2PL model logistics on item 10 with $a = 1,553$, $b = -1,146$, and $c = 0$.

C. Item Analysis model logistics 3PL

The logistic model 3PL involves three parameter that is the difficulty index of item, discrimination power and pseudoguessing. Items analysis results shown in the following table.

Table 4. Difficulty Index of Item

Difficulty Index		
Category	Total	Percentage (%)
Good	22	55
Not Good	18	45
Total	40	100

Based on the table results of analysis using model of logistics 3PL, there are 22 good item or about 55% from all items that has examined based on the item difficulty index. While the rest categorized items is not good as much as 18 item or about 45%.

Table 5. Discrimination Power of Item

Discrimination Power		
Category	Total	Percentage (%)
Good	16	40
Not Good	24	60
Total	40	100

Based on table known analysis results with 3PL model logistics there are good discrimination power of item as much as 16 item or 40% from all items that has examined, and then not good discrimination power of item as much as 24 item or 60%.

Table 6. Pseudoguessing of Item

Pseudoguessing		
Category	Total	Percentage (%)
Good	23	57
Not Good	17	43
Total	40	100

Analysis results on the table 3PI logistics model get pseudoguessing value of item. They are 23 good item pseudoguessing or approximately 57% and not good pseudoguessing value 17 item or 43%. Pseudoguessing value criteria use between 0,0 until 0,2. Estimation of pseudoguessing are frequently close to $1/m$ for multiple choice ability test items with m options. In attitude measurement, nonzero value of c may be infrequent and, except for general response tendencies, their meaning is not obvious. Estimating c is usually difficult because they are very few individuals with extremely low value of θ whose item response can be used to help determine c . The value of pseudoguessing is not well categorized in over the limit predetermined criteria. Item characteristics curve for item 11 shown in the following curve.

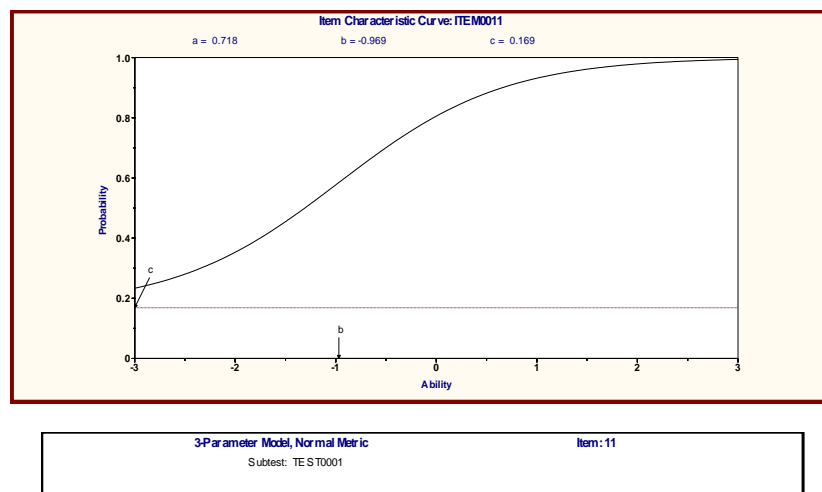


Figure 3. Curve shown an ICC 2PL model logistics on item 11 with $a = 0,718$, $b = 0,969$, and $c = 0,169$.

Table 7. Resume of Good Item and Not Good Item

Category	1 PL	2 PL	3 PL
Good	21	28	23
Not Good	19	12	17

Table shown there are 21 good item on the 1PL logistics model, 28 good item on the 2PL logistics model, and 23 good item on the 3PL logistics model. Then, there are not good item on 1PL as much as 19 item, 12 not good item on 2PL logistics model, and 17 not good item on 3PL logistics model. In these models the most good point obtained in the 2PL logistics model and least in the 1PL logistics model.

D. Determine Fit Model

There are two ways that are used to determine the fit model of analysis to be used, they are determine the fit model in statistics and plot the item characteristics curve. On the selection of fit

model statistically, third model made based on the values of Chi-squared. The fit model can be found by comparing the value of Chi-squared tables with a certain degree of freedom. Items are said to be fit with model if the value of Chi-squared count does not exceed the value of Chi-squared tables. Fit can be known from the value of the probability (of significance, sig). If the value $\text{sig} < \alpha$ items did not fit with the model. Analysis results item characteristics with BILOG MG on high stakes testing senior high school on chemistry 2014/2015 in Jambi province there are 40 item that analyzed by 1PL logistics model shown in the following table.

Table 8. Resume of Item Fit Model 1PL

Item	b	X^2 observed	dk	X^2	Status
28	1.75	10	8	15.507	Fit
29	-1.766	273.2	9	16.919	Not Fit
30	-1.974	288.1	9	16.919	Not Fit
31	-3.241	243.8	9	16.919	Not Fit
32	-3.113	200.2	9	16.919	Not Fit
33	2.833	91.3	7	14.067	Not Fit
34	-3.201	246	9	16.919	Not Fit
35	1.438	50.6	8	15.507	Not Fit
36	1.936	118	8	15.507	Not Fit
37	1.397	230.8	9	16.919	Not Fit
38	-2.196	166	9	16.919	Not Fit
39	-2.227	254.6	9	16.919	Not Fit
40	2.473	123.9	7	14.067	Not Fit

Table shows item fit with model. The table shows only a portion of the item that fit and not fit with compare the value of Chi-square table with values of Chi-square of analysis results with certain degrees of freedom. Items categorized fit with model if the value of Chi-square observed no larger than the value of Chi-square tables. On the 1PL logistics model only one item that fit with the model, the value of Chi-square is 10. That value is not higher than the value of Chi-square tables 15.507 on degrees of freedom (df) = 8. Analysis results item characteristics 2PL logistics model shown in the following table.

Table 9. Resume of Item Fit Model 2PL

Item	a	b	X^2 observed	dk	X^2	Status
1	1.556	-1.977	13.9	7	14.067	Fit
10	1.553	-1.146	29.9	7	14.067	Not Fit
11	1.236	-1.295	13.3	7	14.067	Fit
15	1.08	-2.5	1	7	14.067	Fit
16	0.096	0.268	88	8	15.507	Not Fit
17	0.049	2.469	205.3	8	15.507	Not Fit
18	0.068	1.25	13.27	8	15.507	Fit
19	0.049	0.868	201.7	8	15.507	Not Fit
20	0.036	2.128	264.5	8	15.507	Not Fit
21	1.065	-1.821	10.9	7	14.067	Fit
22	0.071	0.414	131.5	8	15.507	Not Fit
23	0.235	0.45	12.5	8	15.507	Fit
31	1.041	-0.461	9.9	7	14.067	Fit
39	1.367	-1.151	11.5	7	14.067	Fit
40	0.103	-0.868	8.28	8	15.507	Fit

Table shows item fit with model. On the 2PL model there are 9 item fit model and 31 item not fit model. The value of Chi-square is not higher than the value of Chi-squared tables on certain degrees of freedom (df) on item 1, 11, 15, 18, 21, 23, 31, 39, and item 40.

Analysis results item characteristics 3PL logistics model shown in the following table.

Table 10. Resume of Item Fit Model 3PL

Item	a	b	c	X^2 observed	dk	X^2	STATUS
14	0.421	13.304	0.011	43.6	6	12.592	Not Fit
15	1.214	-1.766	0.39	2.5	7	14.067	Fit
16	0.106	29.566	0.007	128	8	15.507	Not Fit
17	0.05	36.943	0.008	291	8	15.507	Not Fit
18	1.549	0.383	0.206	3	9	16.919	Fit
19	0.052	34.302	0.01	262.2	8	15.507	Not Fit
20	0.486	-0.929	0.237	14.02	7	14.067	Fit
21	1.329	-1.034	0.359	9.7	7	14.067	Fit
22	0.678	-0.915	0.214	6.5	8	15.507	Fit
23	1.543	0.338	0.214	3.1	8	15.507	Fit
24	2.354	-0.108	0.12	72.9	7	14.067	Not Fit

Table shows item fit with model. On the 3PL model there are 6 item fit model and 34 item not fit model. The value of Chi-square is not higher than the value of Chi-squared tables on certain degrees of freedom (df) on item 15, 18, 20, 21, 22, and item 23.

On a data, logistics model have item fit model most widely chosen as models for data analysis (Heri Retnawati, 2014). On the test analysis high stakes testing senior high school focus on chemistry 2014/2015 in Jambi province there are 40 items that analyzed by 1PL model, 2PL model, and 3PL model shown in the following table.

Table 11. Resume of Item Fit Model on 1PL, 2PL, and 3PL Model

Model	1PL	2PL	3PL
Fit Model	1	9	6
Not Fit Model	39	31	34

Analysis results based on table shown the model that result item fit model most widely in 2PL model. Its mean 2PL model is the fit model for analysis items characteristics high stakes testing senior high school focus on chemistry 2014/2015 in Jambi province. Analysis results indicate that the most items fit models is 9 item on the 2PL model caused responden that use to analysis more than 200 responden. It is 710 respondents that use to analysis. This is suitable with Heri Retnawati's (2014) opinion that describe growing number of respondents who used the greater the calculate value of chi-square, so the greater the opportunity to reject the hypothesis item fit analyzed by a model.

IV. CONCLUSIONS AND SUGGESTIONS

A. Conclusion

Logistics model on item response theory consists of 1PL model logistics, 2PL, and 3PL. Analysis of the results obtained that on the 1PL model obtained 21 good item and 19 is not good item but only 1 item that fit with the 1PL logistic model, 2PL model obtained 28 good item and 12 is not good item and 9 item fit the 2PL logistics model. On the 3PL model obtained 23 good item and 17 is not good item and 6 item fit the 2PL logistics model. Based on fit item to model logistics, so the item that fit to model is 2PL with fit item is 9 item. This research can concluded the fit model that used to analysis test item high stakes testing in chemistry 2014/2015 in Jambi province.

B. Suggestion

Should the analysis of item characteristics must be adjusted to fit model that is used so that the required information based on item characteristics can be obtained with the right. With the accuracy the results of the analysis of good item characteristics, so it can be retrieved is a good instrument for measuring students ability, so that student learning achievement is getting tracked his progress.

REFERENCES

- [1] Allen, M. J. & Yen, W. M. 1979. *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole Publishing Company
- [2] Djemari Mardapi. 2008. *Teknik Penyusunan Instrumen Tes & Nontes*. Yogyakarta: Mitra Cendekia
- [3] Djemari Mardapi. 2012. *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: Nuha Litera
- [4] Ebel, R.L. & Frisbie, D.A. (1986). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

-
- [5] Hambleton, R.K. dan Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
 - [6] Hamblenton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.
 - [7] Heri Retnawati. 2014. *Teori Respons Butir dan Penerapannya*. Yogyakarta: Nuha Medika
 - [8] Hou, L., Torre, J. D., & Nandakumar, R. 2014. *Differential Item Functioning Assessment in Cognitive Diagnostic Modeling: Application of the Wald Test to Investigate DIF in the DINA Model*. *Journal of Educational Measurement*. Spring 2014, Vol. 51, No. 1, pp. 980125.
 - [9] Saifuddin Azwar. 2012. *Reliabilitas dan Validitas Edisi 4*. Yogyakarta: Pustaka Belajar

Validity and reliability examination of indicators development materials instruction at Elementary School base on Curriculum 2013

Rochmiyati

Lecture at FKIP, University of Lampung
rochmiyatiwardi@yahoo.co.id

Abstract - Examination of Validity and reliability this research to know about validity and reliability of indicators development materials instruction at elementary school base on Curriculum 2013. This study is research and development indicators development materials instruction with scientific, authentic, and contextual approach for development cognitive aspect, affective aspect and psychomotor aspect. Subject of this research are the teachers of Elementary School Pilot Project Implementation Curriculum 2013 at Bandar Lampung. Validated indicators development of materials instruction have been done by expert of Curriculum, Technology of Instruction, Social Science Elementary School, Science Elementary School and Mathematic Elementary School. Examination validity of this step by qualitative analysis and examination of reliability by the General Multifaceted Model. Validity examination indicators development materials instruction by teachers are qualitative analysis, Item Total Statistic and examination of reliability by Cronbach's Alpha. The results of research and development indicators development materials instruction at elementary school base on Curriculum 2013, are as follows; (1) Specification of indicators development materials instruction at elementary school base on Curriculum 2013 with scientific, authentic, and contextual approach for development cognitive aspect, affective aspect and psychomotor aspect. (2) The result of validation by experts show that indicators development materials instruction at elementary school base on Curriculum 2013 with scientific, authentic, and contextual approach for development cognitive aspect, affective aspect and psychomotor aspect are very good and the result analysis by statistic program shows that are reliable. (3) The result of validation by teachers' show that are good, and the result analysis by statistic program shows that are valid and reliable. Indicators have been designed with scientific, authentic, and contextual approach for development cognitive aspect; affective aspect and psychomotor aspect are feasible for development materials instruction at elementary school base on Curriculum 2013.

Keywords: Curriculum 2013, development model, elementary school, indicators, materials instruction, validity examination, reliability examination.

I. PRELIMINARY

Research as a scientific work requires the principle of accountable, transparent and objective. An issue that must be faced in the research activities of education and other social sciences is how to obtain accurate data, careful and objective. This is important because the results can be trusted if the information obtained is reliable. Objectivity accurate data on research using quantitative approach can only be credible, through the measurement procedure. The conceptual framework developed by theoretical framework generates variable, sub-variables and operational indicators (observable) as manual data collection can be properly quantified through the measurement process.

The theoretical framework that has been broken down into variable and the indicators, does not guarantee the quality of the data collection instrument, meaning that necessary quality assurance data collection instrument after testing in order to obtain accurate data and reliable. A good instrument that instrument that can provide information that is valid, reliable and objective.

The level of validity and reliability is shown by the level of accuracy and precision instrument. This level of accuracy instrument influences the decision of a particular variable or more variables. Conclusions are based on accurate data generated from measuring instrument which will help for anyone with an interest in research.

If variables revealed by the instrument have not been tested for validity and reliability levels then the conclusions are determined to be not entirely trustworthy. The conclusion to be decided not by the instrument that measured well is misleading. It cannot be accounted for, in the case of a research as a scientific work must be accountable. Conclusion a good research is the conclusion based on the information data collected by the measuring instrument that meets the right, indicated by the level of validity and reliability indicators variables studied.

Based on the above, the research models of development indicators in primary school teaching materials based Curriculum 2013 must be conducted in order to test the validity and reliability of indicators designed reliable accuracy, can be justified and not misleading. The curriculum integrated curriculum implemented in 2013 as an integrated thematic learning. Integrated in the process, the integrated content and integrated in the assessment.

Implementation of Curriculum 2013, particularly in the city of Bandar Lampung known teachers are still experiencing difficulties in the application of integrated learning, teaching materials development and application of authentic assessment, especially on the high grade (Lilik Sabdaningtyas and Rochmiyati, 2013: 44). Integrated learning practices in primary school pilot project in Bandar Lampung totaling 17 primary schools are still experiencing difficulties in the development of teaching materials.

Curriculum 2013 balancing soft skills and hard skills, covering the aspects of competence attitudes, skills and knowledge as well as formulated into four core competencies are the core competencies first about religious values, competence social values, core competencies third of the knowledge and core competencies fourth on skills known as (K1, K2, K3 and K4). The first and second competency is not the materials that needs to be taught in the classroom, but it needs to be invested through habituation or exemplary so that competence is reflected in our behavior in everyday student life. Religious values and social verbally and attitudes reflected in learning implemented in habits and exemplary teachers who are expected to impact the students. Translation of each core competency is poured into the Basic Competence as a basis for developing indicators of learning outcomes.

Based on the study of the development of the model-based teaching materials Curriculum 2013 Primary School, indicators of development of teaching materials (1) The attitude of spiritual (K1) reflects the attitude of the faithful, noble [honest, polite, caring, disciplined, democratic], confident, and responsible interact effectively with the social and natural environment, around the home, school, and playground. (2) Social attitudes (K2) show behavior (honest, discipline, responsibility, caring, polite, friendly environment, mutual aid, cooperation, peace-loving, responsive and pro-active). (3) Knowledge (K3) in a cognitive perspective on factual knowledge, conceptual, procedural in science, technology, art, culture, and humanities. With the insight of humanity, national, state, and civilization-related phenomena and events, as well as applying procedural knowledge in specific areas of study that suit their talents and interests to solve the problem. (4) Skills (K4) with domain impersonation, manipulation, experiences and articulation of an activity or action in the learning process.

Indicators learning competencies adapted to the blueprint development of teaching materials based on the theme of learning, with the integrated model-based transdisciplinary contextual domain in a real-life context and a local-problem in the perspective of cognitive skills. Indicators developed domain-based cognitive, affective and psychomotor. The indicator development regards to cognitive competence of instruction materials is at the step of knowing, understanding, applying, analyzing, evaluating, producing, or creating on knowledge of: fact, concept, and procedure which comprise knowledge's contents such as: technology, art, and humanity. These contents are based on few subjects, those are: religion, civic, Indonesian language, mathematics, science, social, SBDP, and physical education. Curriculum policy as a national policy is centralized, but at the operational level substantive content of the charge is developed in the area, so it is necessary to develop teaching materials in accordance with the conditions of each region, so that the development of teaching materials is done with contextual approach, factual, authentic.

The development of teaching materials on learning scientific approach is appropriate to use development resources that are locality, so that students are on the factual knowledge, there are the real science for the real world (Kate Calvin and Penny Gilmer, 2009). Indicators is developed customized approach to the implementation of the unified scientific thematic learning, authentic, contextual, and good language. The indicators related to instructional materials and media. The description of the indicators compiled in Table 1.

Table 1. Indicators development of instruction materials

APPROACH	Indicators	Instruction materials and media	Source Materials
scientific	Observe Ask Try Process Present Conclude	Image, Text, goods	School environment, student, internet, and other relevant
Authentic	Can be seen Palpable can be felt can be kissed can be heard	Materials / genuine goods Materials / counterfeit / imitation, Pictures	School environment, student, internet, and other relevant
CONTEXTUAL			
a. Religious	Receive Run Appreciate Appreciating practice	Religious doctrine Compulsory Selection Religious symbols real action	Books religious teachings Religious history book Image / World Map, and other relevant
b. Feeling	Excited / happy Sad Empathy / Caring Angry	sing dance declamation Read prose, poetry, rhymes, picture books	history book books story Classic literature books Indonesia Daily news, and other relevant
c. cognitive	Knowing Understand Apply analyzing Evaluate Create/ Make	Number Arithmetic operations etc. Features Equation Difference Sequence (chain sequence)	Field Studies in elementary school books, maps, pictures of heroes, crop-plants, and other relevant
d. psychomotor	Imitation, Manipulation, Articulated experience.	Movement / Sports Painting / drawing Sew play music	Books, natural environment
LANGUAGE (presentation/the use of language in various activities of teachers and students daily)	Alphabet Standard Structure	Letters Word Syllables Sentence Paragraph type the word The use of the word The use of the word make sentences	Areas of study books And other books relevant

II. RESEARCH METHOD

This study tested the validity and reliability of indicators development of instruction materials on curriculum-based integrated learning 2013. The first data source, the results of the validation curriculum specialists, learning and 6 teachers of elementary learning analyzed using by the

General Multifaceted Model. The second data source, the validation results elementary school teachers who numbered 30 people were analyzed qualitative analysis, Item Total Statistics and examination of reliability by Cronbach's Alpha.

III. RESULT AND DISCUSSION

A. Qualitative analysis

Validation of the indicators of teaching materials is conducted to obtain a good indicator of instruction materials, can be used, meet the criteria of a good indicator of teaching materials, valid and reliable. In the first development stage, all of the design of the device models of teaching materials validated by two experts, and seven teachers.

Assessment experts only deliver an opinion and suggestions, which states that the development of instruction materials has been good indicator, meet the development in accordance with the syllabus and integrated thematic learning model. Rate teachers as practitioners use a questionnaire assessing the design of teaching materials developed, and suggested that the implementation time adapted to local conditions.

Analysis of the results of the evaluation format on all components of the indicator materials submitted model adapt Saifuddin Azwar (2005: 108), which calculates the average value of all components that are assessed on the sheet format validation ratings given by the validator using a scale of four. Qualification standards validation results of this study using a scale of five categories in order to obtain more detailed data. To what extent the results of the indicator materials validator, the criteria based, and the category of description quality indicators materials instruction is display on table 2.

The range of scores in the instrument 0 to 4
 Minimum score = 0,
 The maximum score = 4
 Range feasibility five categories.
 So Range Score = $\frac{4-0}{5} = 0.8$

The range criteria of description quality indicators materials instruction development is;

Table 2. The category of indicators materials instruction

Score range	Category
3.21 – 4	Very good
2.41 – 3.20	Good
1.61 – 2.40	Enough
0.81 – 1.60	Less
0 - 0.80	Bad

Description calculation of result assessment about indicators materials instruction they are very good. Trend of score of the result is homogeny there is trend of assessment the same and consistent. Description calculation of the result assessment six (6) validator, show that on table 3.

Table 3. Description the result of validity of materials instruction indicators

No.	Implementation	Average score the result of assessment rater, at						Average	Category
		1	2	3	4	5	6		
1.	Development at topic	4	4	4	3.9	4	3.9	3.96	Very good
2.	Stages of materials instruction development	3.5	3.5	3.8	3.8	3.9	3.6	3.70	Very good
3.	Development materials instruction	3.9	3.5	3.7	3.9	3.8	3.7	3.75	Very good
4.	Example development materials instruction	3.9	3.4	3.9	3.8	3.6	3.6	3.70	Very good

Recurse: Analysis data by rater

Trend of data the result teachers assessment are homogeny at for all design model materials instruction. Show that balancing contribution of every component materials instruction indicators. This can be known at present comparison from teachers assessment when implementation

indicators development materials instruction. Description the result of teachers' assessment at table 4.

Table 4. The result of teachers' assessment about implementation indicators Development materials instruction

Implementation	Approach			
	Scientific	Authentic	Contextual	Language
Development at topic	24,0 %	23,3%	24,8 %	24, %
Stages of materials instruction development	26,0 %	25,9 %	26,2 %	27,1 %
Development materials instruction	23,0 %	24,1 %	24,6 %	23,5 %
Example development materials instruction	27,0%	26,7 %	24, 4 %	25,4 %
Total	100 %	100 %	100 %	100 %

Recourse: Analysis data teachers' assessment

Every indicator is implementation follow as (1) Development at topic, (2) Stages of materials instruction development, (3) Development materials instruction, and (4) Example development materials instruction. All of indicators materials instruction can be stated easily matched to know and to implement at development of materials instruction.

B. Quantitative Analysis

The validity examination of development materials instructor indicators by average measure and computer program and coefficient correlation examination by standard critical point examination statistic >0.204 (Maruyama, 1998: 213), or >0.30 (William R. Dellon & Mathew Goldstein, 1984 : 69) it is valid. The computer calculation result of average measure from statistic is 0.664 (>0.204 and >0.30). By methodological result assessment by teachers for all indicators of materials instruction are correct, believed or valid.

Reliability examination of assessment rater (teachers) at indicators materials instruction by analysis of the general multifaceted model. Result reliability examination about implementation indicators at development topic are 0.63, implementation indicators at stage of development materials instruction are 0.67, implementation indicators in the unity materials instruction are 0.64, implementation indicators at example development materials instruction are 0.74. Result the all reliability examination by statistic program are >0.60 . This result shows that implementation indicators at development materials instruction by teachers' assessment can be reliable, consistence at assessment, so indicators at development materials instruction are reliable.

The result of the tryout of 30 teachers is heterogeneous. The data has been analyzed by validity examination Item Total Statistic (ITS) at SPSS Program. The result of validity criteria is >0.30 at significant 5% is valid. Examination of reliability about indicators development of materials instruction by Alpha Cronbach's with criteria at > 0.60 . The result of teachers' assessment is on the table 4.

Table 5. Result of Validity and Reliability Examination Indicators of Materials Instruction

No.	Implementation	Validity coefficients $>0,30$	Stated	Reliability coefficient s $>0,60$	Stated
1	Development at topic	0.684	Valid	0.656	Reliable
2	Stages of materials instruction development	0.574	Valid	0.654	Reliable
3	Development materials instruction	0.496	Valid	0.629	Reliable
4	Example development materials instruction	0.655	Valid	0.674	Reliable

Recourse: Tryout analysis data at 30 teachers

The result of development materials instruction indicators by scientific approach very exact because of using the source of local development, so the student can study on factual knowledge there are the real science for the real world. The development indicators adapted at thematic instruction implementation with scientific approach, authentic approach, contextual approach, and good language. That's Indicators are related with materials and media of instruction.

IV. CONCLUSION

The results of research and development indicators development materials instruction at elementary school base on Curriculum 2013, are as follows; (1) Specification of indicators development materials instruction at elementary school base on Curriculum 2013 with scientific, authentic, and contextual approach for development cognitive aspect, affective aspect and psychomotor aspect. (2) The result of validation by experts show that indicators development materials instruction at elementary school base on Curriculum 2013 with scientific, authentic, and contextual approach for development cognitive aspect, affective aspect and psychomotor aspect are good and the result analysis by statistic program shows that are reliable. (3) The result of validation by teachers' show that are good and the result analysis by statistic program shows that are valid and reliable. Indicators have been designed with scientific, authentic, and contextual approach for development cognitive aspect; affective aspect and psychomotor aspect are feasible for development materials instruction at elementary school base on Curriculum 2013.

REFERENCE

- [1] Calvin, Kate & Penny J. Gilmer (Eds.). (2009). Real science for the real world. Doing, learning & teaching. Florida: PACE, Florida Department of Education.
- [2] Cronbach, L.J. & Meehl, P.E
- [3] (1995) Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302
- [4] Dillon, William R. and Mathew Goldstein. (1984). Multivariate analysis methods and application. New York: John Wiley & Sons.
- [5] Maruyama, Geoffrey M. (1998). Basic of structural equation modeling. London: Sage Publication.
- [6] Kementerian Pendidikan dan Kebudayaan (2013). Naskah Akademik Pengembangan Kurikulum 2013. Jakarta: Nopember, 2012
- [7] Lilik Sabdaningtyas dan Rochmiyati. (2013). *Need Assessment* pada Implementasi Kurikulum 2013 SD di Kota Bandar Lampung. *Laporan Hasil Penelitian*. Bandar Lampung: Lemlit Universitas Lampung
- [8] Saifuddin Azwar. (2000). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar

Analisis Item Information Function on the Test of Mathematics

Rukli

Muhammadiyah University of Makassar

Hp. 085299135009

Email: a_rukli@yahoo.co.id

Abstract - This paper aims to reveal: 1) the items fitted the three parameter logistic model; 2) the item information function on the parameters of discrimination the test of mathematics; 3) the item information function on the difficulty level of the test of mathematics; 4) the item information function on the guessing chance level of the test of mathematics. The research data using examinee responses on national final examination of the math. Measurement model using three-parameter logistic model. Analysis of the response data using ASCAL program. The results showed that: 1) there are about 25 of the 35 items fitted the three parameter logistic model; 2) the function of each information about the smaller 1 except item 15; 3) the value of the function of discrimination parameter information the highest and lowest, respectively 816,226 and 1,026; 4) value function parameter information difficulty level is above 100 by 46%, while below 100 by 54%; 5) value function parameter information guessing chance level are in the top 50 as much as 20%, while under 50 as much as 80%.

Keywords: *parameter, item, information, logistic*

I. INTRODUCTION

System of entrance exams and final exams at all levels of education are still testing both the shape and dynamics of the implementation according to community needs. This regard, would require a data and processes in order to maximize results in accordance reference environment. Data, processes, and the environment is a factor reference inseparable from the quality of the output of a system including the system of evaluation.

Now, the evaluation system in Indonesia of college entrance examination system can be divided into three namely SNMPTN, SBMPTN, and Independent. Entrance exam of basic education and secondary education until now has not been of special concern to the government except for the final exam. The exam has undergone a metamorphosis, but sometimes only a change of form, but the essential problem is sometimes secondary. In fact, these changes are sometimes too abstract and theoretical political over curriculum changes.

Curriculum changes inseparable variation example Ebtanas final exam and the UN in order to test the system more credible, relevant and contemporary. The third can be used as a reference rate the system a good test. Credible and relevant to the present forecast accuracy while associated with the speed and quality of information the test results to the user.

Linkages credible and relevant, the core evaluation system lies in the prediction problem. Furthermore, the prediction accuracy of estimating the related problem of latent abilities, which revolves around the problem of how do so predictions are not far away or right on the measuring object. Sumadi Suryabrata (1989) suggested that the problem of prediction generally revolve around the issue of how pressing kemelesetan become smaller.

The ability of the test examinees unique and latent that assessment not free from error. The ability to be in a continuous variable that needs to be defined operationally. Response test examinees misses the most extreme enough mistake to mistake normal estimates. It concerns the coefficient of reliability of the test.

Function test information in item response theory has a similarity coefficient reliabilitas tests. However, the reliability of the test has limitations compared to the test information function.

Samejima (Hulin et. al., 1983) suggests reliability is a dead concept theoretically because it differs from one group against another, generalizations are limited and very narrow. Function test information is an accumulation function item information. Accumulation is not found in reliability because there is no information about the reliability of the item. It shows one disadvantage compared to classical test theory item response theory.

Item information functions play a role in the development and selection of appropriate test items informative ability of the test taker. That is, with the test item information, it can be seen how much the items to uncover the optimal parameters of ability and capability where the item could be used because the estimator.

Furthermore, the value of the item information functions should not be constant on the proficiency scale. Information on the value of the ability depends on the number and attributes of the items used to assess ability. For example, the low value of the function information concerning the ability of the test taker at a high skill level given the item easily, but the item will elaborate or to inform a number of things substantial at low skill levels. How big are the different power parameters, the parameters of difficulties, and parameter guesses opportunities simultaneously affect the value of the function of each item of information about the particular characteristics of the parameters on the math test item nasional exam? Characteristics of test items require in-depth study of descriptive and third parameters of test items.

A. Three Parameter Logistic Model

Three- parameter logistic model in IRT has some logistics model that is one parameter logistic model, two- parameter logistic model, and a three-parameter logistic model. Furthermore, it has developed a four parameter logistic model where the fourth parameter is the parameter of time but has not been implemented optimally (Hamblenton, 1989). Three-parameter logistic model has three parameters, namely the different power parameters item, item difficulty levels, and opportunities guesses items. These three parameters are called independent variables. These three parameters either individually or simultaneously affect the ability parameters as the dependent variable. Equation three- parameter logistic model as follows.

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp(Da_i(\theta - b_i))}{1 + \exp(Da_i(\theta - b_i))} \dots\dots\dots (1)$$

Different power items are variables in the three parameter logistic model. These variables affect the variable capability. In the item characteristic curve, different power expressed by the tendency of a tangent curve at the inflection point of the curve, ie the inclination offensive line curve at the inflection point. The greater the value the greater the inclination to form different power curve is getting steeper. In theory, the difference between the power value range of negative infinity to positive infinity, but in general the practice of different power values are ranging from 0 to 2.

The level of difficulty of items to note the abscissa at an inflection point where the item characteristic curve inflection point is a point on the curve with the ordinate $(1 + c_i)/2$. In theory, the range of item difficulty level between negative infinity to positive infinity but in practice, in general, the level of difficulty ranges from negative three to positive three. Difficulty level value of zero to the right more difficult, otherwise the value of the difficulty level from zero to the left is getting easier.

Opportunities guesses items expressed high ordinate at the point of intersection between the axis of the $P_i(\theta)$ item characteristic curve. These parameters are called pseudo-chance level, which is to claim much chance the correct answer by the examinees enabled an item is lower than the level of difficulty of the items.

$P_i(\theta)$ is an opportunity to test examinees θ answered correctly item-level capabilities to-i. D is a constant scaling factor with a value of 1.7.

B. Function information item

Three- parameter logistic model derived, obtained an information function Information function according to three parameters logistic model can be formulated as follows.

$$I_i(\theta) = D^2 a_i^2 \left(\frac{Q(\theta)}{P(\theta)} \right) \frac{(P(\theta) - c_i)^2}{(1 - c_i)^2} \dots\dots\dots (2)$$

Based on the equation 2, Hambleton and Swaminathan (1985) suggests the nature of the function information item, ie the value of the information function maximum is reached when θ it is slightly larger in value than the value of b, the value of the information function increase if the a value is growing, and the value of the information function increase if the value of c become smaller .

The function of the maximum information is obtained when $\theta = \theta_i$ where

$$\theta_i = b_i + \frac{1}{Da_i} \ln \frac{1 + \sqrt{1 + 8c_i}}{2} \dots\dots\dots (3)$$

Maximum function item information is an indication that the item matches the ability of the test taker. The way it is often used to establish adaptive tests. According to Hambleton, et al. (1991) and Lord (1980) adaptive test is a test of the level of difficulty of matching items with the level of ability of the test taker. Nevertheless, the model test as many constrained in terms of the network without wires so much in the testing phase in addition to understanding the test still lay so socialization is very important to the user (Rukli, 2010 and 2012).

Furthermore, the function item information item is a strength measuring the ability of the test taker. It plays a role in assessing the condition of an item is functioning optimally or not the maximum. This becomes further reference to the item worthy or not worthy entry in the test mainly tests nationwide.

II. RESEARCH METHODS

A. Types of research

This research uses descriptive exploratory study. The goal is that the characteristics or properties of one or more variables can be described. Characteristic parameters of the items comprising different power item, item difficulty levels, and opportunities guesses item will be described. Furthermore, these characteristics be considered to determine the function of the information.

B. Object of research

The research object is an item of national exams in mathematics consisting of 3 (three) package which package 1 from 8 rayon , package 2 from the 7 rayon , and package 3 from 8 package.

C. Item analysis

Analysis items using ASCAL to produce an estimate item parameters, namely the different power parameters of items (a) , the parameter level of difficulty of items (b), and the parameters opportunities guesses item (c) and Chi-squared statistic for each item. Estimated parameters of items will be stopped when the difference between two consecutive estimates of less than 0001 and said to be estimated accurately item parameter estimates between iterations so-called convergent estimates. Generally accurate assessment takes 5 to 10 rounds, but this study uses the number of iterations as many as 20 rounds.

The results of this analysis also result in a match against the item response models. Furthermore, the calculation of each parameter item information functions using Visual Basic.

D. Function Item Information

Calculation function information item using the item parameter different power parameters, level of difficulty, and opportunities guesses. The function calculation procedure has several stages of the test examinees response data to making programs with the following details: (1) Data UN test examinees' responses obtained from the database in the area of the object of study. (2) Data response cleared of disability data. (3) The data processing using ASCAL program. (4) Examination of model fit every item on the response data. (5) Examination matches the second stage until all items meet models. (6) 6. Tabulates and mendeskripsikan each parameter item. (7) Make a program to compute the function parameter information item.

III. RESEARCH RESULT

A. Estimated Parameter Item

UN test item data analysis showed round stop on lap 15 of the 20 rounds are maximized, meaning that at the round reached a level of accuracy estimates the highest among the existing lap . Round stop on lap 15 shown in Figure 1 wherein the difference in lap 14 and lap 15 smaller 0.001.

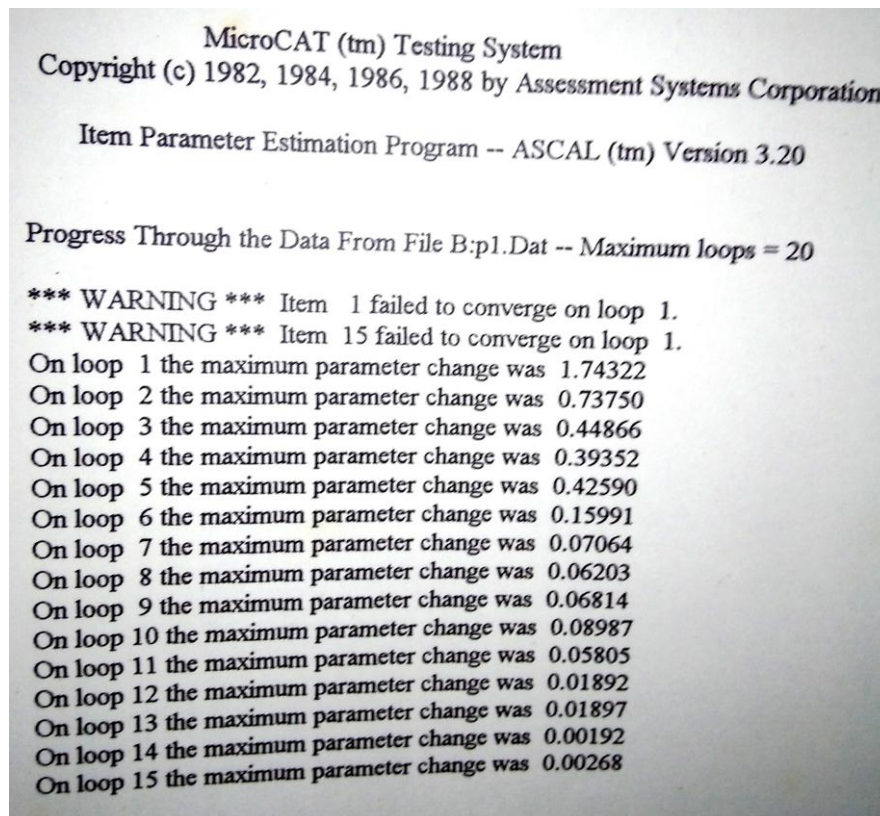


Figure 1. The Print Screen Total Estimated Stop Round

Parameter estimation different power item, item difficulty levels, and opportunities guesses items on lap 15 can be seen in Table 1.

Table 1. Item Parameter Estimates Three Parameter Logistic Model

Item number	Parameter Estimates			Item number	Parameter Estimates		
	(a)	(b)	(c)		(a)	(b)	(c)
1	1.110	3.000	0.200	19	1.008	1.862	0.170
2	0.583	2.700	0.010	20	0.720	0.146	0.000
3	0.691	0.674	0.000	21	0.682	1.467	0.100
4	0.695	0.590	0.070	22	0.497	1.373	0.210
5	0.681	0.550	0.000	23	0.793	1.463	0.150
6	0.400	0.613	0.360	24	0.888	2.267	0.110
7	0.474	0.916	0.170	25	0.659	2.105	0.100
8	0.467	1.582	0.080	26	0.778	2.296	0.150
9	0.774	0.445	0.000	27	0.558	-0.050	0.020
10	0.762	-0.044	0.010	28	0.587	1.446	0.050
11	0.950	1.759	0.210	29	0.692	2.777	0.270
12	0.618	1.687	0.110	30	0.466	1.425	0.020
13	0.4000	0.272	0.270	31	0.895	3.000	2.290
14	1.103	2.745	0.250	32	0.747	2.400	0.350
15	1.571	2.723	0.260	33	0.472	2.936	0.280
16	0.664	1.791	0.230	34	0.416	2.225	0.030
17	1.055	1.027	0.070	35	0.971	2.962	0.120
18	0.944	1.249	0.140				

Table 1 shows that the characteristics of the item has a different power greater than 1 there are five items or 14 %, ie items 1, 14, 15, 17, and 19. The items have different power less than 1 there were 20 or 86 %. Characteristics of the items have no negative different power while the item has a difficulty level equal to 3 No 2 item or 6 %, ie items 1 and 31. Further characteristics of the item has a difficulty level of negative 2 item or 6 %, ie, items 10 and 27. Characteristics items have greater opportunities than 0.300 guess there are 2 or 6 %, ie, items 6 and 32.

B. Item Fit Three Parameter Logistic Model

Compatibility between the data from the work of test takers and the theoretical value of the grain refers to the suitability of the items on the three parameter logistic model. If Chi- Squared value of an item is greater than the critical value in the model, which is 27.6 then the item is not compatible with the model. Conversely, if the value of an item Khi-Square is less than or equal to the critical value, which is 27.6 then the item fits with the model.

Number of items meet the logistic model three parameters as many as 25 of the 35 items, or 71%, namely, 1, 4, 6, 7, 8, 11, 12, 14, 15, 17, 18, 19, 21, 22, 23, 24, 25, 26, 28, 29, 31, 32, 33, 34, and 35. While the item does not meet the model as much as 10 or 29%, ie, items 2, 3, 5, 7, 9, 10, 13, 16, 20, 27, and 30.

Incompatibility between data items with the model likely to be caused in terms of material, construction, and the language of the item. For example in terms of language, lack of clarity of the revelation by the symbol given in the item. Examples on item 2, pernyataan "complement ($A \cap B$)" will confuse the test taker. The statement requires the test taker to understand the meaning of meaning, complement, and sliced, where the two concepts requires different thinking.

The item does not match the model is not working properly to uncover capabilities. An item does not work properly due to the item usual bias or the ability of test takers do not fit the participant's ability to respond to the test items, such as too difficult or too easy.

The items do not match the model can be used for other purposes. Sumadi Suryabrata (1987) revealed that if statistically a characteristic item is not feasible then the item is returned to the test objectives. Furthermore, Hambleton et. al., (1991) reveals that the item does not match the three-parameter logistic model may be matched with two parameters logistic model that only has two different parameters such as power parameters item and item difficulty level. Or may be incompatible with the model of the parameters that only pay attention to the items based on level of difficulty alone.

If you notice an item has a high chance guesses there are 4 items located at the end of the test, namely items 29, 31, 32, and 33. To answer a complicated math test item takes about 4.5 minutes (Sumadi Suryabrata, 1987), while the UN test mathematics SMP 120 minutes with 40 items. Thus, such tests including speed test. According to the research Adkin and Malenkopt in different years (Sumadi Suryabrata, 1987) showed no untoward effect on the estimated parameters of the items to be taken at the end of the test if the processing time narrow test.

Based on this, most likely takers have not tried to do all of the items on the test time is up so that the test taker to guess. That is, the items are less than helpful in revealing the ability of the test taker. However, further analysis of the matter was not done in accordance limitation of this study is a quantitative analysis. Therefore, it can be concluded that the item is an item fits models have different power high, high level of difficulty, and the opportunities and guesses were high among 35 items.

C. Function Parameters Information Item

Item meets the models analyzed to calculate the value of the item information functions. Function information item only performed on the items meet the logistic model. The function of each parameter item of information contained in Table 2.

Table 2. Function Item Information

Item number	Function Item Information			Item number	Function Item Information		
	Ia	Ib	Ic		Ia	Ib	Ic
1	4.474	7.617	1948.87	4	149,089	106.616	1177.80
6	113.891	21.319	785.71	7	201.499	40.215	785.71
8	406.51	37.987	1619.81	11	34.378	31.759	2236.85
12	211.821	40.299	2127.83	14	4.767	8.524	2499.39
15	1.026	10.231	2521.26	17	76.69	125.042	2909.29
18	67.072	69.666	2226.09	19	29.993	30.619	2704.69
21	181.553	51.415	2197.02	22	213.669	29.123	1280.83
23	96.267	45.944	2151.69	24	44.984	20.977	3831.41
25	170.336	26.54	30.11.95	26	60.155	16.647	2927.84
28	329.488	54.877	2279.84	29	35.665	5.692	2182.22
31	6.505	3.005	2299.29	32	26.17	6.607	1901.75
33	152.968	5.44	1973.11	34	816.226	28.174	2968.80
35	10.318	4.438	4319.35				

Table 2 shows that the value of the power parameter information functions depending highest item is item 34, the items have value power parameter information functions depending smallest item is item 15. The items have value parameter information functions difficulty level above 100 there are 11 or 44%, while the items have value parameter information functions difficulty level below 100 there are 14 or 56%. The items have value information function parameter guesses opportunities in the top 50 there are 5 or 20%, while the items have value information function parameter guesses opportunities under 50 there are 20 or 80% .

The correlation coefficient between the estimated value of different power parameters and parameter information function value of -0.623. According to Hambleton and Swaminathan (1985) the tendency of the correlation varies by level of ability. Low skill levels of negative tendencies, whereas a high ability level positive tendency. Furthermore , the correlation coefficient between the estimated value of the parameter level of difficulty and the value of the function parameter information sebesar -0761 and the correlation coefficient between the estimated value parameter guesses opportunities and the value of the information function parameters sebesar -0343

Furthermore, the value of r - table with $n = 25$ obtained 0396 . That is, the correlation between the correlation between the estimated value of the item parameter and function value information item only significant parameter in different power and level of difficulty. Not significant parameter estimated value of the item and the value of the function 's parameter information item guesses opportunities occur because the opportunity is less consistent guesses matching using ad hoc methods. Hulin et. al., (1983) stated that the estimated parameters chance guesses are usually suitable when using a method that involves consideration of researchers in interpreting the pattern of test-takers answer . The way it is difficult to do even can be said can not be done when national exams take place.

IV. CONCLUSION

Items that match the three-parameter logistic model as many as 25 of the 35 items, or 71%. Value function power parameter information depending highest item is item 34, the items have value different power parameter information functions smallest item is item 15. The items have value parameter information functions difficulty level above 100 there are 11 or 44%, while the items have value parameter information functions difficulty level below 100 there are 14 or 56%. The items have value information function parameter guesses opportunities in the top 50 there are 5 or 20%, while the items have value information function parameter guesses opportunities under 50 there are 20 or 80%. The functions of the items are all less infomasi 1, except item 15 has a value greater item information function 1. The low value of the item information functions is different due to the low power, level of difficulty is higher when compared to the ability of the test taker, and guess high enough opportunities.

REFERENCES

- [1] Sumadi Suryabrata, "Seleksi calon mahasiswa baru: perguruan tinggi yang sekarang dan kemungkinan untuk masa yang akan datang," Yogyakarta: Andi Offset Yogyakarta, 1989.
- [2] Hulin, H. L., Drasgow, F., and Parsons, C. K., "Item response theory. Application to psychological measurement," Homewood, IL: Thorsey Professional Series, 1983.

- [3] Hambleton, R. K., "*Principles and selected applications of item response theory*," in R. L. Linn (Eds.), *Educational Measurement*, 1989,. Macmillan, New York, pp.147-200.
- [4] Hambleton, R. K. dan Swaminathan, H., "Item response theory, " Boston, MA: Kluwer, 1985.
- [5] Hambleton, R. K., Swaminathan, H., & Rogers, H. J., "*Fundamentals of item response theory*," CA: Sage, Newbury Park, 1991.
- [6] Lord, F. M., "Application of item response theory to practical testing problems," Hillsdale, NJ: Lowrence Erbaum Associates, 1980.
- [7] Rukli, "Penerapan model logistik satu parameter pada *Computerized Adaptive Testing* Ujian Masuk Perguruan Tinggi;" Prosiding Seminar Nasional Himpunan Evaluasi dan Pendidikan Indonesia, Jakarta. ISBN 978-602-96343-0-3, 2010.
- [8] Rukli, "Simulasi Pola Inisialisasi Kemampuan dengan Metode Penalaran *Fuzzy* Tsukamoto dalam Menidentifikasi *Cheating*," PPS Universitas Negeri Yogyakarta dan HEPI Yogyakarta, Yogyakarta. 978-602-96172-8-3. 2012.
- [9] Sumadi Suryabrata, "Pengembangan tes hasil belajar," Jakarta: Rajawali Press, 1987.

Misuses Cronbach Alpha On Achievement Tests

Satrio Budi Wibowo

FKIP Universitas Muhammadiyah Metro

satrio.budi.wibowo@mail.ugm.ac.id

Abstract - Cronbach alpha (α) formula is widely used in the field of social research and education. The appeal of the formula Cronbach α is, easier to calculate, easy to understand, and is available in several statistical programs. Along with its popularity, also found much less appropriate use of the formula Cronbach α . Inaccuracy use of Cronbach alpha (α) formula can result in low coefficient α obtained. Test makers sometimes hastily concluded that a low coefficient of Cronbach α indicating the low reliability of the test, which is caused by the bad item. Based on this, the test makers finally shed the items from the test, in order to increase the of Cronbach α coefficient. Deleting many items may adversely affect the tests validity. Even though, low coefficient of Cronbach α could be caused by the inaccuracy of the test makers in using Cronbach α formula. This paper will discuss of the inaccuracies in the use of Cronbach α formula on achievement test, which can cause coefficient α becomes underestimate.

Keywords: *Cronbach Alpha, Achievement test, Reliability*

I. INTRODUCTION

The main objective of the development of achievement tests is to promote learning [1]. Evaluation at the end of learning period, through the test, can promote children to study hard in order to get a good test score. Good test scores can be a rewards that can motivate students. Even when score of the test is bad, if understood properly, it still motivates students to learn, in order to avoid a low score test. Therefore, the function of the achievement tests is very important in the learning process.

Achievement test items can be written in various formats, including *multiple choice, matching, true/false, short answer*, and *essay* [14]. These formats vary in their strengths and weaknesses, and no one format is ideal in all circumstances. However, many achievement tests developed using a multiple choice format, because of multiple choice format is more efficient [14].

On achievement tests, the answer of testee will be scored true or false. True or false score are dichotomous (only two categories of scores). Many achievement tests scoring using dichotomous model, perhaps only a few cases where the score are not scored in dichotomy models. For example, achievement tests with three categories score, -1 for the wrong answer, 0 if no answer, and 1 if the answer correctly.

Achievement test results should accurately measure individual differences or achievement at a certain pre-specified mastery level [1]. To accomplish these purposes, a test must be valid and reliable [1]. Validity is addressed when a test plan is formulated to accurately represent the course content and depth of learning achieved in a course [1]. Test results must be reliable or repeatable to be confident that a student's score is a true reflection of an examinee's achievement [1].

When testing the reliability of a test. The makers of the test will be faced with the various formulas for estimating the reliability. Alpha (α) Cronbach formula is most often used by researchers in fields of social and educational [2]; [3]. Web of science reporting, more than 6500 cites that use a Cronbach α formula [2]. One of the reasons why Cronbach α reliability formulas most commonly used is because it's simple and can be found in many statistical computer programs.

However, the alpha (α) Cronbach formula sometimes do not fit if used for estimating the reliability of the specific tests types. On achievement tests that using dichotomous scoring models, using alpha (α) Cronbach formula is not recommended. There is another reliability formula that more suitable for estimating dichotomy scoring test models.

II. CRONBACH ALPHA (A) FORMULA

Scores obtained from the test that measuring non-physical attributes, not necessarily depict the subject actual ability. Scores obtained by subject on a test only observed score that a sum of two part, true score and error score, it can simply be formulated in the following formula [4];

$$X = T + e \quad (1)$$

Where : T= True Score, X= Observed Score, e = Error score.

Various theories in psychometric were developed as an effort to estimate how actually the subject T score, only rely on information from the X score value were obtained from the test. One of the assumptions on classical test theory are the T score is assumed as the mean of the theoretical distribution of X scores that would be found in repeated independent testing of the same person with the same test ($E(X) = T$) [4]. This assumption in practice is imposible to be tested. Various ways are sought to guess the T score by applied the classical test theory assumptions, one of them with estimating the reliability. Estimation of reliability aims to predict how large random error that occur in measurement results, so that we can then determine the measurement accuracy [5]. When the measurement accuracy is well known, further, researchers were able to predict the T score value obtained by the subject.

There are many ways to estimate the test reliability. A test is said to be reliable if the observed score (X) correlates highly with true score (T) [4]. If the correlation is perfect, $eg = 1$, then certainly the error value is 0, or it can be said there is no measurement error at all (see formula 1). Thus, the squared correlation between the observed score and true score can be regarded as a test reliability (ρ_{xt}^2). Squared correlation (ρ_{xt}^2) can be obtained by the following formula [4];

Observed score variance (σ_x^2) by the assumptions of classical test theory is as follows [6];

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 + 2\sigma_{te} = \sigma_t^2 + \sigma_e^2 \quad (2)$$

Covariance between the observed scores and true scores σ_{xt} based on classical test theory is [6];

$$\sigma_{xt} = \sigma_t^2 + \sigma_{te} = \sigma_t^2 \quad (3)$$

The correlation between the observed score and true score can be simplified as follows;

$$\rho_{xt} = \frac{\sigma_{xt}}{\sqrt{(\sigma_t^2 + \sigma_e^2)(\sigma_t^2)}} = \frac{\sigma_t^2}{\sqrt{\sigma_x^2 \sigma_t^2}} = \frac{\sigma_t}{\sigma_x} \quad (4)$$

Then, when squared, will obtain the following formula;

$$\rho_{xt}^2 = \frac{\sigma_t^2}{\sigma_x^2} \quad (5)$$

A test may also be proved reliable if the test observed score (x) correlated with observed scores from the parallel tests (x'), the notation denoted by $\rho_{xx'}$. Due to the correlation between two parallel test score is equal to the square of the correlation between true score and observed score [4]. Thus, the correlation between the two parallel test scores can be represented by the following formula [4];

$$\rho_{xx'} = \rho_{xt}^2 = \frac{\sigma_t^2}{\sigma_x^2} \quad (6)$$

Based on the formula (6), it can be concluded, the greater the value of $\rho_{xx'}$, the greater degree of the confidence in guessing the true score (T) based on the observed scores (X) that obtained from the measurements, due to errors in measurement will be smaller.

Cronbach α reliability coefficient, based on the coefficient developed by Guttman and Kuder Richardson formula [7]. When estimate the reliability, the test will be split into several parts. Item covariance formula was actually first developed by Richardson (1937) [8], who developed a shortened version of the two parts of the Spearman-Brown formula [6]. The test parts denoted by k (if the test is split as much as the existing items, k can also be indicated as the number of

items), while the item variance denoted by σ_i^2 . Item variance in the alpha Cronbach formula is used to replace the true score variance (σ_t^2) that impossible to be diestimesi. Thus the Cronbach α coefficient formula is as follows;

$$\rho_{xx} = \frac{\sigma_t^2}{\sigma_x^2} = \frac{k^2 \frac{\sigma_x^2 - \sum \sigma_i^2}{k(k-1)}}{\sigma_x^2} = \frac{k}{k-1} \frac{\sigma_x^2 - \sum \sigma_i^2}{\sigma_x^2} \quad (7)$$

Cronbach α coefficient representing the estimation ratio of the true score variance with a total score variance, it can be seen in formula (7) [9]. As the two sides of the coin, the coefficient alpha estimates the number of measurement errors, when the measurement error decreases, the coefficient alpha will be increase, and vice versa [10]. For example, if the test has a coefficient alpha = 0.80, it means that there is a variance of 0.36 error (random error) in the score ($0.80 \times 0.80 = 0.64$; 1.00 to $0.64 = 0,36$) [11]. After getting a score of measurement errors then we can estimate how much the actual range of true score values obtained by the subject. However, that the coefficient alpha only revealed the effects of measurement error in the score was observed from a group of subjects, not describe the effects of measurement error in an individual score.

III. MISUSES CRONBACH ALPHA FORMULA ON ACHIEVEMENT TESTS

Cronbach α reliability coefficient used to calculate the test/scale internal consistency when the scores are polytomous, such as "1 to 5" [4] [12]. When Cronbach α used for achievement tests with the dichotomous scoring type, sometimes will produce underestimate reliability score. Several simulations have been carried out by the researchers to test the statement. One of the illustrations can be seen in the following sample data.

Table 1. Example Dicotomous Score Test

Subjek	Item										Total
	1	2	3	4	5	6	7	8	9	10	
A	1	1	1	1	1	1	1	1	1	1	10
B	0	0	1	0	0	0	0	1	1	0	3
C	1	1	1	0	0	1	1	1	0	0	6
D	1	1	1	1	1	1	1	0	1	1	9
E	0	0	0	1	0	0	1	0	0	0	2
F	1	1	1	1	1	1	1	1	1	1	10
G	0	1	0	1	1	1	0	1	1	1	7
H	1	1	0	1	1	1	1	1	1	0	8

Keterangan:

$$\sum \sigma_i^2 = 2,375$$

$$\sigma_x^2 = 9,267$$

$$\sum p(1-p) = 2,078$$

Based on example data from the above, and the application of the formula (7), then the computation obtained as follows.

$$\alpha = \frac{10}{10-1} \frac{9,267 - 2,375}{9,267}$$

$$\alpha = 0,826$$

The coefficient α value will be higher, if the calculations are done using a KR 20 formula. KR 20 formula developed by Kuder and Richardson (1937) [8]. KR20 formula are very similar to the Cronbach α formula, because Cronbach α formula developed derived from KR 20 formula [6]. KR 20 formula can be seen as follows:

$$KR\ 20 = \frac{k}{k-1} \frac{\sigma_x^2 - \sum_{i=1}^N p_i (1-p_i)}{\sigma_x^2} \quad (8)$$

If Cronbach α using the total amount of variance item ($\sum \sigma_i^2$) (see formula 7), the KR 20 uses a score of chances to answer correctly (p) (see formula 8). The data in Table 1, when estimated using the KR 20 formula will be obtained the following results;

$$KR\ 20 = \frac{10}{10-1} \frac{9,267 - 2,078}{9,267}$$

$$KR\ 20 = 0,861$$

Result show increased reliability coefficient ($\alpha = 0.826$, $KR\ 20 = 0.861$) when dicotomous data are estimated using KR 20 formula.

Actually, KR-20 and alpha coefficients are same coefficients [13]. Differences occur in the way these two formulas translate true score variance (σ_t^2). Alpha coefficient translated true score variance as follows;

$$\sigma_t^2 = \frac{k^2 \frac{\sigma_x^2 - \sum \sigma_i^2}{k(k-1)}}{(9)}$$

When in KR 20 coefficient true score variance translated as follows;

$$\sigma_t^2 = \frac{k^2 \frac{\sigma_x^2 - \sum_{i=1}^N p_i(1-p_i)}{k(k-1)}}{(10)}$$

Can be seen from 9 and 10 formula, the difference occurs when the KR 20 uses correct answer opportunities replaced items variance on Cronbach alpha formula. This difference minimize the error on dicotomy score type when using the formula KR 20. Thus, the simulation data obtained variance item score on the formula KR 20 ($\sum p(1-p) = 2,078$), lower than the variance item estimate using Cronbach alpha formula ($\sum \sigma_i^2 = 2,375$).

IV. CONCLUSIONS AND RECOMMENDATIONS

As it is well known KR-20 and alpha coefficients are same coefficients; however, it calls KR-20 when it is used for dichotomously scored items and it calls alpha when it is used for polytomously scored items [13]. Alpha is a general version of the Kuder-Richardson coefficient of equivalence. It is a general version because the Kuder-Richardson coefficient applies only to dichotomous items, whereas alpha applies to any set of items regardless of the response scale [15]. The use of formula Cronbach α on dichotomously scored items can cause underestimate reliability coefficient. So some experts, suggested using a KR 20 or KR 21 when item scores are dichotomy [12].

REFERENCES

- [1] Flateby, T. L. (1996). A guide for writing and improving achievement tests. Tampa: University of South Florida, Office of Evaluation and Testing. Retrieved from <http://archive.aacu.org/meetings/generaleducation/gened2013/documents/Flateby.pdf>
- [2] Stijma, K. (2009). On The Use, Misuse, And The Very Limited Usefulness Of Cronbach's Alpha. *Psychometrika* 74(1), 107-120.
- [3] Seref, T. (2009). Misuses of KR-20 and Cronbach's Alpha Reliability Coefficients. *Education and Science*, 34(152).
- [4] Allen, M. J., & Yen, W. M., (1979). *Introduction to Measurement Theory*. California : Brooks/Cole publishing Company
- [5] Cronbach, L. J. (2004). My current thoughts on Coefficient Alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.
- [6] Revelle, W. (2015). *An Introduction To Psychometric Theory With Applications in R (E-Book, Under Development)*. Diunduh dari <http://www.personality-project.org/r/book/>
- [7] Peterson, R. A. (1994). A meta-analysis of Cronbach's alpha. *Journal of Consumer Research*, 21, 381-391.
- [8] Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2, 151-160.
- [9] Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of mental test scores*. Reading MA: Addison-Wesley.
- [10] Teo, T., & Fan, X. (2009) Coefficient Alpha and Beyond: Issues and Alternatives for Educational Research. *Asia-Pacific Edu Res*, 22(2), 209-213. DOI 10.1007/s40299-013-0075-z
- [11] Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*. 2. 53-55. DOI: 10.5116/ijme.4dfb.8dfd
- [12] McMillan, J. H. (1992). *Educational research: fundamentals for the consumers*. New York: Harper Collins.
- [13] Tan, S. (2009). Misuses of KR-20 and Cronbach's alpha reliability coefficients. *Egitim ve Bilim*, 34(152), 101.

- [14] González, S. A., & Vivó, C. D. (2001). The definition of achievement and the construction of tests for its measurement: A review of the main trends. *Psicológica: Revista de Metodología Y Psicología Experimental*, 22(1), 43-66.
- [15] Cortina, J., M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology* 78(1), 98-104.

ITEM DISCRIMINATION OF TWO TIER TEST ON HYDROLYSIS OF SALT

Sri Yamtinah*, Haryono, Sulistyo Saputro, Bakti Mulyani, Suryadi BU
Chemistry Education Program, Sebelas Maret University
jengtina_sp@yahoo.com

Abstract - The aim of this research was determined item discrimination of two tier test that developed for Hydrolisis of Salt. Hydrolysis of salt is one of the difficult material for senior high school students based on a survey carried out on students and interviews with teachers, so it is important to be measured by using the precise instrument assessment. The conventional multiple choice instrument will not be able to measure precisely because only in a score dichotomy of right and wrong that was not able to observe the process of thinking learners. Two tier test chosen as the appropriate instrument to measure students' abilities. This test consists of two parts, namely a first tier and second tier. Students were required to choose a reason for answers to be chosen in the first tier. Item discrimination was an important criterion in the development of test instruments including the two tier test. Good items were that have discrimination index ≥ 0.3 . By a good discrimination index of the items said to be able to distinguish either the students who has the ability and not in solving problems.

Keywords: *discrimination item, two tier test, hydrolysis of salt*

I. INTRODUCTION

Chemistry is the science that is acquired and developed through experimentation seeking an answer to the question of what, why, and how relevant certain natural phenomena; especially those relating to the composition, structure and properties, transformation, dynamics, and energetics substances. Chemical subjects in high school aims to learn everything about the composition, properties and structure, transformation, dynamics, and energetics of substances and their application to solve everyday problems. Chemistry is a product (chemical knowledge in the form of facts, concepts, principles, laws and theories) the findings of scientists and process (Purba, 2006).

Chemistry is considered as a science that is less interesting and difficult because the chemical closely related to ideas or abstract concepts and vocabulary specific, so where studying chemistry such as learning a new language (Chang, 2005). The weakness of learning process, students are less encouraged to develop the ability to think is a problem faced today. Learning in the classroom more directed to the ability to memorize information without the demand to understand the information, resulting in a smart student would only theoretically not apply.

There are three levels in studying the chemical knowledge that is macroscopic, sub-microscopic, and symbolic (representational) (Barke, 2009:27). The relationship between these levels should be taught explicitly. Relationship between the level can be described as follows:

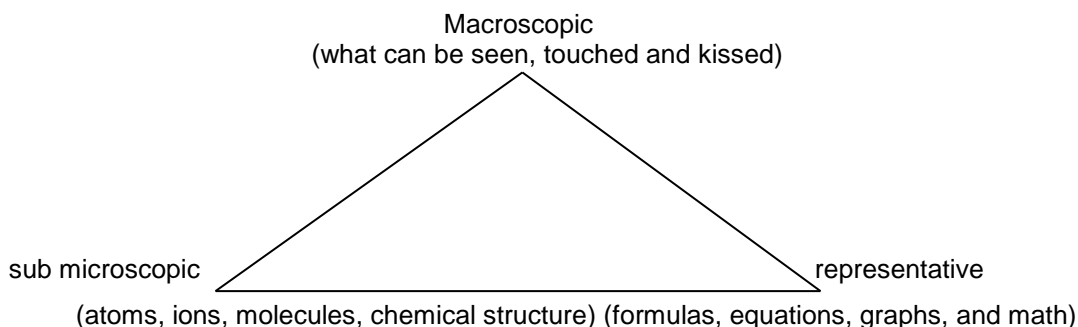


Figure 1. Three Level in Chemical Sciences (Sirhan, 2007)

A macroscopic level where the phenomenon can be observed in the visible. Sub-microscopic level is a phenomenon that can not be observed in the visible. While the symbolic level is the depiction of macroscopic and sub-microscopic level into a symbol. For example, corrosion of the spikes on the macroscopic level the students can see there is brown carats. What happened to the sub microscopic level is a chemical process in which the surface of the iron reacts with oxygen in the air and produces a molecule of iron oxide. To be able to explain the events of oxidation and reduction on the spikes, it can be explained through the concept of iron corrosion equation with symbols, formulas and figures, namely: $4\text{Fe (s)} + 3\text{O}_2 \text{ (g)} \rightarrow 2\text{Fe}_2\text{O}_3 \text{ (s)}$.

In the study of chemistry all of three levels are taught in their entirety. This is why students sometimes find it difficult to learn chemistry, especially to learn in sub-microscopic abstract concepts. Based on the description above, it can be concluded that learning of chemical is a process interaction between learners with their environment to learn everything about a substance that includes the composition, structure and properties, changes energetics involving substances and reasoning skills (Johnstone, 2006).

Basic materials in the chemical is closely related between one material with other material. Although not as tight as in mathematics and physics, in chemistry learning, the order of each concept are concerned. For example before the learners master the concept of salt hydrolysis learners must first master the acid-base balance and pH. Similarly, before the students learn the concepts of chemical bonding in advance learners must master the concepts of atomic structure. Therefore steps of learners in work on the problems of chemical must be ordered according to the hierarchy of existing concepts.

A mandate carried curriculum education in 2013 is about the essence of a scientific approach to learning. The scientific approach intended in the curriculum of 2013 is learning the scientific approach which emphasized using science process skills. Science process skills is an approach, which in practice students are required to be able to apply scientific methods during the learning activities (Semiawan, 1992).

Science process skills are transferable intellectual thought process on a skill. Science process skills are categorized in two things: the basic skills and integrated skill. Basic process skills including observing, inferring, measuring, communicating, classifying, predicting, using the relationship of space and time using numbers. While the integrated process skills include controlling variables, operationally defining, formulating hypotheses, formulate models, interpreting the data and experimenting (Chabalengula, Mumba and Mbewe, 2011).

The success of the learning process can be seen through the assessments conducted on learners. Assessment is an activity taken the decision to determine the things based on good and bad criteria and a qualitative. Assessment attributed with

one's abilities, such as intelligence, skill, accuracy, and others associated with the work or duties. (Putra, S.R, 2012: 14).

Assessment to measure science process skills require precise instruments. One instrument that can be used to measure the knowledge aspect of science process skills is a Two-Tier Multiple Choice (TTMC). TTMC in this study consists of the main problems in the form of multiple choice as first tier and about the reasons that also multiple choice as a second tier.

The rules of development of two-tier multiple-choice diagnostic instrument used in this study is described by Treagust (1988). In this instrument, the first tier of each item consist of a content question of five choices; the second part of each item contain a set of five justifications for the answers to the first part. Included in these justifications are the correct answer and two to five distracters. Distracters are derived from students' alternative explanations gathered from the literature, interviews and free responses. In advocating different assessment procedures to probe students' understanding of scientific concepts, Simpson and Arnold (1982) recommended that information relating to erroneous information held to be true by students should be included in the tests that have distracters.

This line of research in assessment has included the development of multiple choice tests items that have distracters based on students' conceptions. Two-tiered questions have two main benefits over conventional one-tiered questions. The first is a decrease in the measurement error. The second benefit to the two-tiered format is that it allows for the probing of two aspects of the same phenomenon. This allows the probing of the phenomenological domain with the first tier and the conceptual domain with the second.

The item analysis is an important phase in the development of an exam program. In this phase statistical methods are used to identify any test items that are not working well. If an item is too easy, too difficult, failing to show a difference between skilled and unskilled examinees, or even scored incorrectly, an item analysis will reveal it. The two most common statistics reported in an item analysis are the item difficulty, which is a measure of the proportion of examinees who responded to an item correctly, and the item discrimination, which is a measure of how well the item discriminates between examinees who are knowledgeable in the content area and those who are not.

The item discrimination index is a measure of how well an item is able to distinguish between examinees who are knowledgeable and those who are not, or between masters and non-masters. There are actually several ways to compute an item discrimination, but one of the most common is the point-biserial correlation. This statistic looks at the relationship between an examinee's performance on the given item (correct or incorrect) and the examinee's score on the overall test. For an item that is highly discriminating, in general the examinees who responded to the item correctly also did well on the test, while in general the examinees who responded to the item incorrectly also tended to do poorly on the overall test.

The possible range of the discrimination index is -1.0 to 1.0; however, if an item has a discrimination below 0.0, it suggests a problem. When an item is discriminating negatively, overall the most knowledgeable examinees are getting the item wrong and the least knowledgeable examinees are getting the item right. A negative discrimination index may indicate that the item is measuring something other than what the rest of the test is measuring. More often, it is a sign that the item has been mis-keyed.

When interpreting the value of a discrimination it is important to be aware that there is a relationship between an item's difficulty index and its discrimination index. If an item has a very high (or very low) p-value, the potential value of the discrimination index will be much less than if the item has a mid-range p-value. In other words, if an item is either very easy or very hard, it is not likely to be very discriminating. A typical CRT, with many high item p-values, may have most item discriminations in the range of 0.0 to 0.3. A

useful approach when reviewing a set of item discrimination indexes is to also view each item's p-value at the same time. For example, if a given item has a discrimination index below .1, but the item's p-value is greater than .9, you may interpret the item as being easy for almost the entire set of examinees, and probably for that reason not providing much discrimination between high ability and low ability examinees.

In the two-tier items, discrimination index is determined on each tier. Determination of the index of discrimination by using point biserial formulas.

II. RESEARCH METHOD

This research is a descriptive study was part of the research development. The study used an instrument Two-Tier Multiple Choice (TTMC) include individual profile (Yamtinah, et al 2015). Samples were students from the high school category of high, medium and low total 96 students in the city of Surakarta.

III. RESULT AND DISCUSSION

Analysis of the need for the interests of the need for types of instruments that can measure the science process skills obtained as follows:

1. Teacher as much as 94.73% applying the assessment test to measure the ability of learners. Tests conducted in the form of multiple choice and essay.
2. Teacher as much as 94.73% sees the need of an alternative form of testing in addition to multiple-choice test and essay test, which is able to measure the Science Process Skills (KPS) students effectively.
3. Teacher as much as 52.63% chose salt hydrolysis materials as one of the materials is quite difficult in learning chemistry grade XI.

Based on the results of questionnaire analysis of the needs of teachers can be concluded that the system of assessment used by teachers in the form of multiple choice and description, in this case the form of multiple choice questions and a description of each have weaknesses and have not been able to sharpen understanding of the concept of students as a whole and has not been able to measure the skills process science students. So according to the teachers need to be holding another alternative form of matter that can measure students' science process skills. Developed form of matter is a kind of two tier. Questions two tier developed to have five options on the first tier and the third option on the second tier. One example is:

Pada reaksi kimia berikut yang merupakan reaksi hidrolisis garam adalah....

- A. 50 ml CH_3COOH 0,1 M dan 100 ml NaOH 0,1 M
- B. 50 ml HNO_3 0,5 M dan 150 ml NaOH 0,3 M
- C. 100 ml CH_3COOH 0,1 M dan 50 ml KOH 0,1 M
- D. 100 ml HCl 0,1 M dan 100 ml NaOH 0,2 M
- E. 50 ml HF 0,1 M dan 100 ml KOH 0,05 M

Alasan dari jawaban saya adalah....

- A. Hidrolisis terjadi apabila larutan asam dan basa yang direaksikan habis bereaksi.
- B. Hidrolisis terjadi apabila larutan asam lemah / basa lemah bersisa
- C. Hidrolisis terjadi apabila asam kuat direaksikan dengan basa kuat.

In the two-tier issues developed scoring is not dichotomy that true = 1 and false = 0 but is made in the form of graded as follows

Student Answer		Score
First tier	Second tier	
correct	correct	3
correct	Wrong	2
wrong	Correct	1
wrong / no answer	wrong / no answer	0

Moreover, the responses were analysed using Iteman software program. While analyzing these data, the scoring was considered using Graded Response Model (Yamtinah, 2015). The reliability of the instrument was established by a Cronbach alpha coefficient of 0.865 for first tier and 0.870 for second tier. If reliability coefficient is over 0.70-0.80, instrument is reliable (Özgüven, 1998) and our value was also consistent with other two-tier tests (e.g. Chandrasegaran, Treagust & Mocerino, 2007).

Item analysis carried out using software Iteman get a difficulty index and the index of discrimination as follows:

No	Main field test	Item Analysis		No	Main field test	Item Analysis	
		Ind Dif	Ind Disc			Ind Dif	Ind Disc
1	First tier	0.750	0.569	11	First tier	0.375	0.398
	Second tier	0.781	0.392		Second tier	0.604	0.774
2	First tier	0.708	0.547	12	First tier	0.573	0.703
	Second tier	0.458	0.547		Second tier	0.688	0.688
3	First tier	0.760	0.450	13	First tier	0.323	0.463
	Second tier	0.625	0.366		Second tier	0.677	0.624
4	First tier	0.458	0.569	14	First tier	0.385	0.680
	Second tier	0.844	0.295		Second tier	0.573	0.604
5	First tier	0.823	0.363	15	First tier	0.719	0.473
	Second tier	0.917	0.263		Second tier	0.667	0.691
6	First tier	0.646	0.263	16	First tier	0.656	0.827
	Second tier	0.833	0.144		Second tier	0.688	0.763
7	First tier	0.771	0.366	17	First tier	0.531	0.556
	Second tier	0.885	0.212		Second tier	0.521	0.623
8	First tier	0.281	0.339	18	First tier	0.604	0.771
	Second tier	0.552	0.241		Second tier	0.615	0.784
9	First tier	0.531	0.730	19	First tier	0.594	0.528
	Second tier	0.604	0.702		Second tier	0.625	0.628
10	First tier	0.281	0.459	20	First tier	0.510	0.603
	Second tier	0.552	0.456		Second tier	0.823	0.527

In the second tier item question no 6 shows that the index of difficulty 0.833 meaningful is an easy matter but has an index of discrimination 0.144 meaningful discrimination power is low (less able to differentiate students who are capable and incapable). At this item the revision on the option answer choices. Item discrimination indexes below 0.19 should be eliminated or completely revised (Crocker & Algina, 1986). The others item generally have a good an index of discrimination.

IV. CONCLUSION

A total of 20 items in a two tier test on Hydrolysis Salt material, resulting have an index of discrimination good enough so that no one else aborted items. Some of the items that have low index of discrimination revised option answer choices.

REFERENCES

- [1] Adodo, S. O. (2013). Effects of Two-Tier Multiple Choice Diagnostic Assessment items on Students' Learning Outcome in Basic Science Technology. Ondo State: Academic Journal of Interdisciplinary Studies by MCSER-CEMAS-Sapienza University of Rome. 2 (2)
- [2] Barke, H.D, Al Hazari & Yitbarek S. (2009). Misconceptions in Chemistry, Addressing Perception in Chemical Education, German: Springer-Verlag Berlin Heidelberg
- [3] Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. Chemistry Education Research and Practice 8(3), 293-307
- [4] Chabalengula, V.M., Mumba, F. & Mbewe, S. (2012) How Pre-service Teachers' Understand and Perform Science Process Skills. Eurasia Journal of Mathematics, Science & Technology Education, 8(3), 167-176
- [5] Chang, R (2003). Kimia Dasar:Konsep-konsep Inti. Jakarta: Erlangga
- [6] Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Florida: Holt, Rinehart and Winston INC

-
- [7] Johnstone, A. H, (2000). Theaching of Chemistry-Logical or Pshycological. Chemistry education : Research and Practice in Europe, 1(1), 9-15
 - [8] Purba, Michael. (2007). Kimia jilid 2B untuk SMA Kelas XI. Jakarta: Erlangga
 - [9] Putra, S.R. (2012). Desain Evaluasi Belajar Berbasis Kinerja. Jember: Diva Press
 - [10] Semiawan, Conny dkk. (1992). Pendekatan Ketrampilan Proses. Jakarta: Gramedia
 - [11] Sirhan, G. (2007). Learning Difficulties in Chemistry: An Overview. Journal of Turkish Science Education, 4(2), 5-8
 - [12] Yamtinah, Haryono dan Martini K.S (2014). Profil Individu Peserta Didik Pelengkap Tes Jenis Testlet Sebagai Alternatif Pendeteksi Kesulitan Belajar Kimia. Jurnal Profesi Pendidik. 1(1), 1-10

An Analysis of Test Quality by Using ITEMAN

¹Tia Nur Istianah, ²Desrin Lebagi

¹English Education Department, Sebelas Maret University

²English Education Department, Sebelas Maret University

tianur.07@gmail.com

Abstract—This study used descriptive qualitative research in which the researchers intended to analyze and describe the quality of the tests by using ITEMAN (Item Test Analysis). The analysis involves the difficulty level, the discrimination index, distracters index, test validity and the reliability of the test. This program is used not only to analyze an English test, but also to analyze other tests. Therefore, in this study, we analyze two different tests namely, English Test and Natural Science test. The result shows that one of the tests is more qualified than another one by considering the criteria of a good test item. Based on the research findings, by utilizing ITEMAN, we can identify the quality of test items simultaneously and give value judgment of each item whether it is a good item or not. The result of this study is expected will bring beneficial information for the teachers in analyzing the quality of their own test.

Keywords: *Test Quality, ITEMAN*

I. INTRODUCTION

Test is one of the elements of curriculum that must be conducted in teaching learning process. When the teacher wants to give the test, it should be appropriate to the learning goal or objective. According to Crobach in Indira (2008:4), a test is a systematic procedure for observing a person's latent trait and describing it with the aid of numeral scale or a category system. While Suryabrata in Pasoloran (2009:6) argues a test as a question that should be answered and or the commands which must be done based on how the examinee answer the question or do the commands, examiner takes the conclusion and compare with another standard. This highlights that test is an instrument or procedure used to measure students' comprehension and ability. Another theory also support this statement that the term test refers to a variety of ways of collecting information on learner's ability or achievement (Dorobat, 2007:2).

The test can give many benefits not only for the teacher, students, but also for the progress of the teaching learning process. However, the problem that usually faced by the teacher is the way they measure the quality of their test. Teacher get difficulty in measuring the test manually, because it is time consuming. On the other hand, it is a must for the teacher to construct a qualified test in order to be able to measure students' ability and comprehension. Therefore, to know whether the test is qualified or not, teacher needs to construct and analyzed the test based on the criteria of a good test. A good test has criterias that include its difficulty level, discrimination index, distracters index, validity and reliability.

A good test includes some difficult items to test the higher students and some easy items to give lower students a chance. An item can be categorized as a good test if they are in the moderate criteria. It means the item is not too easy and not too difficult. Besides, an achievement test with superior discriminating power is needed. To obtain this, it is necessary to avoid the use of test items which widely varying difficulty level. Distracters are meant to test students' understanding of the questions. Principally, the distracters are designed to approach the similarity with answer keys in terms of length, word class, and so other nature of the answer key. Tinambunan (1988:11) argues that validity refers to the extent to which the results of an evaluation procedure serve the particular uses for which they are intended. Thus, the validity of the test is the extent to which the test measures what is intended to measure. James and Sally (2006:179) define reliability as the consistency of measurement the extent to which the results are similar over different forms of the same instrument or occasions.

Those criteria of a good test can be analyzed by using ITEMAN that stands for Item Test Analysis. ITEMAN is a device or program that help the teacher to analyze item and test easily. This program is included in one package with MicroCat developed by Assessment System

Corporation began in 1982 and revised in 1984, 1986, 1988, and 1993: from version 2.00 to version 3.50. Statistics of test item in ITEMAN includes Number or Scale-Item that refers to the number of item. Prop-Correct refers to difficulty index of the item. Point Biser interprets discrimination index of the item. While, Prop-Endorsing explains the distracters of each item. In addition, this program can analyze the test comprises 10 scales (subtest) and provide information about the validity and reliability of the test.

II. RESULTS AND FINDINGS

A. The Analysis of Tests

After inputting and analyzing data on ITEMAN program, the result of ITEMAN analysis provides important information to interpret the quality of those different tests including level of difficulty, power of discrimination, distracter index, reliability and validity. The interpretation is taken by comparing the result of ITEMAN analysis and several criteria of a good test.

1. Difficulty Level

After analyzing the data using ITEMAN, the levels of difficulty of those test can be interpreted as follow: The analysis result of English test shows that of 30, there are 12 items including in easy category, 13 items including in moderate category, and 5 items including in difficult category. Ideally, the percentage of difficulty level distribution in a test should be 20% items in easy category, 40% in moderate category, and 20% in difficult category. By looking at this result, it can be concluded that this whole test does not have good level of difficulty because there are 40% items categorized in easy category, 43% items categorized in moderate category, and only 16,66% items categorized in difficult category.

On the other hand, the analysis of Natural Science test shows that of 40 items, there are 11 (27.5%) items including in easy category, 21 (52.5%) items including in moderate category, and 8 (20%) items including in difficult category. Regarding to the ideal percentage of difficulty level distribution, it can be concluded that this test has good distribution of difficulty level.

In determining the category of each item, the researchers follow those following criteria:

0.00 – 0.30	= difficult
0.31 – 0.70	= moderate
0.71 – 1.00	= easy

Below are the examples of items categorized in easy, moderate, and difficult categories.

2. Item in easy category

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
19	0-19	0.760	0.562	0.409	A	0.120	-0.228	-0.140	
					B	0.760	0.562	0.409	*
					C	0.000	-9.000	-9.000	
					D	0.120	-0.646	-0.397	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

This item is categorized in easy category because the prop. Correct is 0.760.

3. Item in moderate category

6	0-6	0.440	0.481	0.382	A	0.440	0.481	0.382	*
					B	0.040	0.523	0.230	
					C	0.360	-0.440	-0.343	
					D	0.160	-0.288	-0.191	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

This item is categorized in moderate category because the prop. Correct is 0.440.

4. Item in difficult category

9	0-9	0.240	0.244	0.177	A	0.600	-0.303	-0.239	
					B	0.040	0.362	0.159	
					C	0.240	0.244	0.177	*
					D	0.080	-0.049	-0.027	
					E	0.040	0.200	0.088	
					Other	0.000	-9.000	-9.000	

This item is categorized in difficult category because the prop. Correct is 0.440.

B. Discrimination Power

In relation to the result of ITEMAN analysis, the power of discrimination of those test can be interpreted as follow: The result shows that the English test has 6 items which need to be retained, 9 which subject to revision, and 15 item that must be omitted. The percentages of discrimination power in this test are as follow: 20% items have good power of discrimination, 30% items have marginal power of discrimination, and 50% items have poor power of discrimination and those should be omitted.

Furthermore, in Natural Science Test, of 40 items there are 16 (40%) items which need to be retained, 10 (25%) items which need to be revised, and 14 items that need to be omitted. The percentages of discrimination power in this test are 40% items have good discrimination, 25% items including have marginal discrimination power, and 35% items have poor discrimination power. In determining the power of discrimination of each item, the researchers adopt those following criteria:

0.40 and up	: Very good items and to be retained
0.30 – 0.39	: Good, but the items possibly subject to revision
0.20 – 0.29	: Marginal items, but the items usually need to revision
Below 0.19	: Poor items, and to be rejected

Below are the Examples of items that have very good, good, marginal, and poor power of discrimination shown on the result of ITEMAN analysis:

1. Item with very good power of discrimination

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				Key
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	
25	0-25	0.240	0.691	0.503	A	0.360	-0.403	-0.314	
					B	0.320	-0.042	-0.032	
					C	0.040	0.039	0.017	
					D	0.240	0.691	0.503	*
					E	0.040	-0.607	-0.267	
					Other	0.000	-9.000	-9.000	

This item has very good power of discrimination shown on the point biser about 0.503.

2. Item with good power of discrimination

20	0-20	0.640	0.403	0.314	A	0.320	-0.314	-0.241	
					B	0.640	0.403	0.314	*
					C	0.000	-9.000	-9.000	
					D	0.000	-9.000	-9.000	
					E	0.040	-0.446	-0.196	
					Other	0.000	-9.000	-9.000	

This item has good power of discrimination shown on the point biser about 0.503.

3. Item with marginal power of discrimination

16	0-16	0.320	0.308	0.236	A	0.560	-0.128	-0.102	
					B	0.120	-0.298	-0.183	
					C	0.000	-9.000	-9.000	
					D	0.000	-9.000	-9.000	
					E	0.320	0.308	0.236	*
					Other	0.000	-9.000	-9.000	

This item has marginal power of discrimination shown on the point biser about 0.236.

4. Item with poor power of discrimination

15	0-15	0.680	0.042	0.032	A	0.000	-9.000	-9.000	
					B	0.000	-9.000	-9.000	
					C	0.000	-9.000	-9.000	
					D	0.680	0.042	0.032	*
					E	0.320	-0.042	-0.032	
					Other	0.000	-9.000	-9.000	

This item has poor power of discrimination shown on the point biser about 0.032.

C. Distracter Index

In interpreting the quality of each distracter in each item shown on the result of ITEMAN analysis, the researchers pay attention on prop. Endorsing and the second point biser on the table. If the point biser of the distracter is in negative, it indicates that the distracter works well. However, if the point biser is in positive, it does not work well. Besides, the quality of each distracter in each item can be seen on the prop. Endorsing on the table by considering on the following criteria:

76% - 125%	= very good dis. index
51% - 75% or 126% - 150%	= good dis. index
26% - 50% or 151% - 175%	= poor dis. index
0% - 25% or 176% - 200%	= bad dis. index
> 200%	= very bad dis. Index

Below are the representations of distracters quality shown on the result of ITEMAN analysis:

Item Statistics					Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
7	0-7	0.720	0.055	0.041	A	0.040	-0.446	-0.196	
					B	0.200	0.010	0.007	
		CHECK THE KEY			C	0.720	0.055	0.041	*
	C was specified, E works better				D	0.000	-9.000	-9.000	
					E	0.040	0.200	0.088	?
					Other	0.000	-9.000	-9.000	

On that table, the answer key is C. Regarding to the prop. Endorsing and point biser of each distracter, the quality of each distracter can be interpreted as follow:

- Distracter A : its Prop.Endorsing is 0,040 or 4%. It means that the quality of this distracter is bad, but it has negative value of biser indicating that it functions well (0,040>0,025).
- Distracter B : its Prop.Endorsing is 0,200 or 20%. The quality of this distracter is bad because most of students consider it as the right answer. Besides, the positive value in point biser indicates that it does not function well.
- Distracter D : its Prop.Endorsing is 0,000 or 0%. It is bad distracter index, and does not function well (0,000<0,025).
- Distracter E : its Prop.Endorsing is 0,040 or 4%. It means that the quality of this distracter is bad. Also, it has positive value of biser and does not function

well meaning that many students in upper class choose this option as the correct answer.

D. Validity

An item can be categorized as valid item if the biser value is 0.3(coefficient of validity) or higher than 0.3 (biser \geq 0.3). In accordance with the result of ITEMAN analysis, it is shown that of 30 items in English test, only 13 items that are valid. Furthermore, of 40 items in Natural Science test, 26 items are valid. It based on the biser value shown on the table. Below is the example of item validity based on the result of ITEMAN analysis.

Seq. No.	Scale -Item	Item Statistics			Alt.	Alternative Statistics			Key
		Prop. Correct	Biser.	Point Biser.		Prop. Endorsing	Biser.	Point Biser.	
13	0-13	0.920	0.142	0.078	A	0.040	-0.284	-0.125	*
					B	0.040	0.039	0.017	
					C	0.920	0.142	0.078	
					D	0.000	-9.000	-9.000	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

By looking at the table, it can be interpreted that this item is not valid because the biser value is 0.142 that is lower than 0.3 (coefficient of validity).

E. Reliability

Referring to the result of ITEMAN analysis, the reliability of each test can be identified by looking at the Alpha index written on statistic scale and comparing the index to the coefficient of reliability. If the Alpha is ≥ 0.7 (coefficient of reliability), the test is reliable. If the Alpha is < 0.7 , the test is unreliable. Below are the statistic scales of those tests:

Statistic Scales

English Test

Scale Statistics	

Scale:	0

N of Items	30
N of Examinees	25
Mean	17.760
Variance	8.262
Std. Dev.	2.874
Skew	-0.072
Kurtosis	-0.069
Minimum	11.000
Maximum	24.000
Median	18.000
Alpha	0.387
SEM	2.250
Mean P	0.592
Mean Item-Tot.	0.230
Mean Biserial	0.324

Natural Science Test

Scale Statistics	

Scale:	0

N of Items	40
N of Examinees	20
Mean	22.500
Variance	25.250
Std. Dev.	5.025
Skew	0.823
Kurtosis	-0.399
Minimum	17.000
Maximum	34.000
Median	20.000
Alpha	0.708
SEM	2.717
Mean P	0.563
Mean Item-Tot.	0.266
Mean Biserial	0.342

By looking at the scales and considering the criteria of a good test, it can be concluded that the English test is unreliable, meanwhile the Natural Science test is reliable.

III. RESEARCH METHODOLOGY

The method used in this research is descriptive study with the aim to describe the real phenomenon either a natural phenomenon or unreal phenomenon. Bogdan and Taylor in Moloeng (2000:3) state that the results of this research are in qualitative data, in which the data are collected in the form of written and oral words that are descriptively analyzed. Research samples are English and non-English test that are chosen by a purposive sampling procedure. The purposive sampling is meant to get the homogeneity of the data which are colored by given considerable context. Since this study in the form of descriptive research, the researchers used

document analysis technique in the form of tests analysis. The data aimed to support the success of the research. After collecting the data, analyzing the data becomes the final step to find out the result of this study. The data were analyzed by using Miles and Huberman's model. Miles and Huberman suggested that qualitative data analysis consists of three procedures, they are as follows:

1. *Data reduction*, to reduce process of mass data that are obtained. The data are the test package that consists of some stems. The stems of the test are reduced and organised. The data reduced by avoiding essay or short answer test because this research only analyzed multiple choice test, and the data are organized and analyzed by using ITEMAN.
2. *Data display*, to draw conclusion from the mass of data. After reducing and organizing the mass data by using ITEMAN program, the data results are displayed in the form of table and also brief explanation.
3. *Conclusion drawing or verification*, researcher's conclusion regarding to the result of the study. The conclusion can be verified through some relevant references or theories about criteria of a good test.

IV. CONCLUSION

Considering the results and findings, it can be concluded that several items of those different tests still need to be revised and even omitted. Since, the distribution of difficulty level of the test items, discrimination power, and distracter index of each item is not ideally distributed. Thus, it is suggested for the teachers to analyze their test using ITEMAN program, so that they will be able to construct a qualified test as a tool to measure students' ability and comprehension.

REFERENCES

- [1] Indira. W, Analyzing Discrimination Value of Subjective Test Designed by English Teacher of the First Year Students of SMP N 10 Palu, Thesis, Palu: Universitas Tadulako, 2008.
- [2] Dorobat. Dumitru, The Methodology of Evaluation and Testing. The Ministry of Education and Research Project, ISBN 978-973-0-04814-8, 2007.
- [3] James, H. McMillan and Sally.S, Seventh Edition Research in Education Evidence based Inquiry, Pearson, 2006.
- [4] Moloeng. Lexy J, Metodologi Penelitian Kualitatif, Bandung:Rosdakarya, 2000.
- [5] Pasoloran. H, Analyzing Discrimination Value of the Subjective Test made by an English Teacher of the Tenth Grade of SMA N 5 Palu, Thesis, Palu: Universitas Tadulako, 2009.
- [6] Tinambunan. W, Evaluation of Students Achievement, Jakarta:Departement Pendidikan dan Kebudayaan, 1988.

AN ANALYSIS OF PERSON FIT USING RASCH MODEL

Yessica Mega Aprita¹⁾, Yolandaru Septiana²⁾

Department of Educational Research and Evaluation, Yogyakarta State University

¹ yessiyessicaa@gmail.com, ² yolandaruseptiana@gmail.com

ABSTRACT - This research is a quantitative descriptive research which aims to analyze and describe the response patterns of final examination in the district of Gunungkidul (Special District of Yogyakarta) based on person fit methods. Analyses of the data in this research are based on students' responses in the final examination of Accounting subject of the Grade XI SMK Financial Accounting Expertise Program. Data are obtained from the students' answer sheets for 103 responses which each consists of 40 items multiple choice questions. The results of the research using Winsteps Software indicates that the Person Reliability in the Accounting Final Examination is 0.81 while the Item Reliability is 0.93. The score of Alpha Cronbach is 0.82. Person Measure 0.94 logit shows that the average ability of students is above the item, which means that the response pattern of the students is fit. Additional informations can be obtained are the value of logit person, for the Student-1 with logit +4.07 indicates the person with the highest ability (able to do almost all of the questions); Student-61 with logit -0.98 shows the students with the lowest ability (lowest in answering the questions correctly). Misfit occurs for Students-1 and Student-31 Based on the research results, it can be concluded that most of the participants in Accounting Final Examination have the response patterns that fit, so it is assumed that in this Final Examination of Accounting is not found inappropriateness on student response patterns.

Keywords: *Final Examination, Rasch Model, Person Fit*

I. INTRODUCTION

Evaluation of learning in the end of the semester is a routine activity held every year at various levels of school, especially vocational schools. Basically the final exams held to measure the achievement of students' competence undertaken by the education unit in recognition of the students' achievement. Ideally, test execution does not simply monitor how much ability of participants in answering the items as much as possible, but an attempt to understand the participants' behavior in a test execution. It is important to obtain more varied information in decision-making. Understanding the participants' behavior in an implementation of the test is also useful to know the character of the participants, even it can become a reference in evaluating the implementation and supervision of examinations. To overcome the problems, we implement the person fit method in this research.

Person-fit is a statistical index that can be used as a direct measure of the measurement accuracy and identification of response patterns peculiar based measurement model used. Individuals test scores who have low person-fit value potential for not accurately measure their pure ability (Bambang, 2013: 70). Person fit shows how far the individual response to the instrument according to the model. Thus, an individual has a high value of a fit person if his response to the instrument in according to the model.

There are many things that caused the individual responses not fit with the model, the cause is not due ability / trait being measured. The reason is because of other things outside ability/trait measured, such as drowsiness, carelessly, or they do not understand the instructions. These things cause the individual response to instrument to be weird, odd, or anomalies. In English called the aberrant response.

Based on the background of the problems, then the problem formulation of this research is: "how is the response patterns of participants final exams at SMK Muhammadiyah Wonosari Financial Accounting Skills Program based on fit person method?". Then, the purpose of the research is to analyze and describe patterns of participants' responses final exams at SMK

Muhammadiyah Wonosari Financial Accounting Skills Program based on fit person method. Furthermore, the expected benefits of the implementation of this research for teachers, as information regarding the technique identifies the examinee ability through the person fit method, as well as the Education Office in District of Gunungkidul as a basis for policy decisions in the implementation of the final examinations, especially in SMK Financial Accounting Skill Program in the district of Gunungkidul.

II. RESEARCH METHODS

Generally, this research uses a quantitative approach. The research was conducted at SMK Muhammadiyah Wonosari on first Semester Academic Year of 2015/2016. This research focused on the test of Accounting Subject Final Examination, the total is 40 items. The form of test of Accounting Subject Final Examination is objective multiple choice test with the option of five possible answers developed by KKG Team appointed by the authorities. The data collection technique in this research is documentation. In this research, the total examinee's answer sheets are 103 sheets. The research data that has been captured and then analyzed by a fit person method using Rasch models to analyze the students' ability. To support the implementation of the analysis, we use Winsteps software.

III. RESULTS AND DISCUSSION

This research analyzed the ability of students with a Rasch model using the data of Accounting Final Examination at SMK Muhammadiyah Wonosari Accounting Expertise Program to see the results of students' learning for one semester. Data are obtained from the students' answer sheets for 103 responses which each consist of 40 items multiple choice questions. Below is the data obtained:

Table 1. Students' Response on Accounting Final Examination

STUDENT_1	111111110111111111111111111111111111
STUDENT_2	111011111111110101111111101111110010111
STUDENT_3	1110110111110101110111011101111111011110
STUDENT_4	111111111100110111011101110111011101010
STUDENT_5	01111101111111010111011001011000100101010
STUDENT_6	1111100111100110000100101001010110111110
STUDENT_7	1111111111111111111111111111111111111110
STUDENT_8	1110101001100100000011100101111101100000
STUDENT_9	1011100011000101000010001001100001110111
STUDENT_10	1110000011101111110111111000110110111100
STUDENT_11	0111000111100101011111100111110111010011
STUDENT_12	1110110111101101011001010011010001111110
STUDENT_13	1111010111101111111111111011110111111111
STUDENT_14	11101011101101011111111110111110101111010
STUDENT_15	0001111111100111110111101001011101100100
STUDENT_16	11110001111011011111111111101111111011010
STUDENT_17	1111100100110110011110010001111111001001
STUDENT_18	0110110111110100111110001101111100000100
STUDENT_19	11111011111111011100100010110101111010100
STUDENT_20	111101111110111011111101111101111110101111
STUDENT_21	111111011010111011111111101100111101001111
STUDENT_22	1110100111100100011011101010110111010100
STUDENT_23	1011100010001101010100001011110010000101
STUDENT_24	10111101111001111111111111011111101111110
STUDENT_25	01111001111011111111011111101111111111110
STUDENT_26	1111100111101001011010001011111011110111
STUDENT_27	11110011111111111010011000111111111101110
STUDENT_28	1111010101101101001111100111110011001001
STUDENT_29	1111111111110111111111111111111111111111

STUDENT_30	0001111111011101011101101100111001111000
STUDENT_31	11111111111110111111111101111111111111
STUDENT_32	1101100000100101101110001001011101010101
STUDENT_33	0110100111101111111111111001011111011100
STUDENT_34	0011100111100001011111101001111010001010
STUDENT_35	0100110101111110110111111111110101111010
STUDENT_36	1101100111101101010010100011010000011100
STUDENT_37	1110111111111111111111101111111111111110
STUDENT_38	0011001111100100111110110000111100111110
STUDENT_39	101001111111100010111100101111101111100
STUDENT_40	11111101111011101111101110111111111110110
STUDENT_41	1001001000100101100000001000010010100010
STUDENT_42	0010000111101100010000110000010001100010
STUDENT_43	1110111111110100101111100111111111111101
STUDENT_44	11111111111011111110111110111111111001010
STUDENT_45	1110100101101101101111001000111111101100
STUDENT_46	110010010010010111111100001011111101101
STUDENT_47	0111111111101111111100010111111111111100
STUDENT_48	0000101100100001111100011001010100100011
STUDENT_49	000100011110011101001010101010101111100
STUDENT_50	1111100111110111111101110001111001011100
STUDENT_51	1011110101101111111110110011111001111000
STUDENT_52	1101100000100101101110001001011101010101
STUDENT_53	0110100111101111111111111001011111011100
STUDENT_54	0011100111100001011111101001111010001010
STUDENT_55	0100110101111110110111111111111010111010
STUDENT_56	1101100111101101010010100011010000011100
STUDENT_57	1110111111111111111111101111111111111110
STUDENT_58	0011001111100100111110110000111100111110
STUDENT_59	101001111111100010111100101111101111100
STUDENT_60	11111101111011101111101110111111111110110
STUDENT_61	1001001000100101100000001000010010100010
STUDENT_62	0010000111101100010000110000010001100010
STUDENT_63	1110111111110101111111100111111111111101
STUDENT_64	11111111111011111110111110111111111001010
STUDENT_65	1110100101101101101111001000111111101100
STUDENT_66	1100100100100101111111100001011111101101
STUDENT_67	0111111111101111111100010111111111111100
STUDENT_68	0000101100100001111100011001010100100011
STUDENT_69	000100011110011101001010101010101111100
STUDENT_70	1111100111110111111101110001111001011100
STUDENT_71	1011110101101111111110110011111001111000
STUDENT_72	0011001010101101110110011001010101010110
STUDENT_73	0111101111101111101110101101110100111100
STUDENT_74	0110101111101111111111100101111111001111
STUDENT_75	0100110111110110001010000101101011011111
STUDENT_76	1111101111111011101101100111111111111101
STUDENT_77	11101011111110111011110111111111111011101
STUDENT_78	0111111111101101111111111111111111111110
STUDENT_79	0001110010111101101110100011111100101110
STUDENT_80	0111110011111111111110110101111111101100

STUDENT_81	011110011111010011111000111111100001100
STUDENT_82	01111011010101111101101101101101100110
STUDENT_83	0110101111101101111110000101011111100110
STUDENT_84	01111110111110110111101111111111111110
STUDENT_85	01101011011111011011111010111111111111
STUDENT_86	01100000011111111011011100111110011100
STUDENT_87	0110001011000101110011100110100001000001
STUDENT_88	01111101111110111111010111111110001100
STUDENT_89	0110101100101101011000101001111000000110
STUDENT_90	010011110110110110000010100010100100101
STUDENT_91	1100000111100100101100101101110111000101
STUDENT_92	111111001010010111111111101110111101111
STUDENT_93	0000001111110100010010000010101101011100
STUDENT_94	110100011110010111001001010101010100000
STUDENT_95	1101111111011011111010111011111111010001
STUDENT_96	1111101111101111111111011010111111101110
STUDENT_97	0110100101111111111101011001111100000011
STUDENT_98	1101010111111011011001000110110001000
STUDENT_99	111101011110110011011001010110110001000
STUDENT_100	0111000111101100100000000100100111100101
STUDENT_101	111101011110110011011010010110110001100
STUDENT_102	11110101111011011111000010111111101000
STUDENT_103	0011111100100100000001000000111000001001

The data on the table above analyzed by the Rasch model using Winsteps Software. Here is the analysis output of Winsteps:

SUMMARY OF 103 MEASURED Person

	TOTAL			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	26.4	40.0	.94	.41	1.00	.1	.98	.1
S.D.	6.3	.0	1.00	.12	.12	.8	.35	.9
MAX.	39.0	40.0	4.07	1.03	1.34	2.4	3.56	3.1
MIN.	12.0	40.0	-.98	.35	.78	-1.6	.28	-1.2
REAL RMSE	.43	TRUE SD	.90	SEPARATION	2.09	Person	RELIABILITY	.81
MODEL RMSE	.42	TRUE SD	.91	SEPARATION	2.15	Person	RELIABILITY	.82
S.E. OF Person	MEAN = .10							

Person RAW SCORE-TO-MEASURE CORRELATION = .97

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .82

SUMMARY OF 40 MEASURED Item

	TOTAL			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	67.9	103.0	.00	.25	1.00	.1	.98	.0
S.D.	16.5	.0	1.01	.07	.11	1.1	.28	1.2
MAX.	99.0	103.0	1.75	.52	1.34	3.0	2.16	3.6
MIN.	34.0	103.0	-2.65	.21	.87	-1.5	.56	-1.4
REAL RMSE	.27	TRUE SD	.97	SEPARATION	3.67	Item	RELIABILITY	.93
MODEL RMSE	.26	TRUE SD	.97	SEPARATION	3.73	Item	RELIABILITY	.93
S.E. OF Item	MEAN = .16							

Figure 1. The Output of Summary Statistics

Another important informations that can be obtained based on the Summary Statistics table above are *Person Reliability* and *Item Reliability*, the value of alpha Cronbach and *Person Measure*. It can be seen that the Person Reliability in the Accounting Final Examination is 0.81 while the Item Reliability is 0.93. The amount of Alpha Cronbach is 0.82. Person Measure 0.94 logit shows that the average ability of students is above the item, which means that the response pattern of the students is fit.

The measure column shows the logit value for each item Column measure shows the logit value for each of the items are sorted from highest to lowest. For item number 40 is 1.75 logit shows that these item is the hardest item, while item number 11 is the easiest item amounted to -0,98 logit. To check the items that do not fit (outliers or misfits) can be done by reviewing the following matters.

1. The value of Outfit Mean Square (MNSQ) accepted is in the range of 0,5 – 1,5
2. The value of Outfit Z-Standard (ZSTD) accepted is in the range of -2,0 – 2,0
3. The value of Point Measure Correlation (Pt Mean Corr) in the range of 0,4 to 0,85

Here is the output of person statistics.

Person STATISTICS: MEASURE ORDER

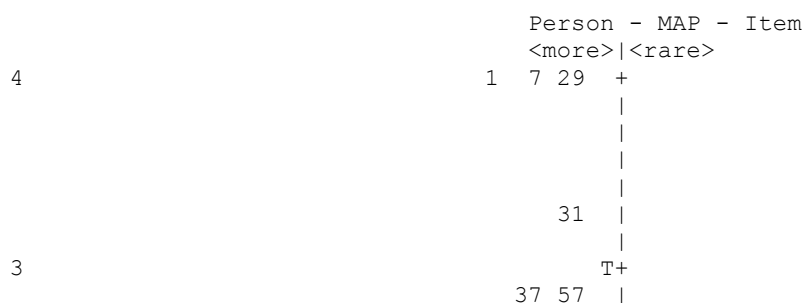
ENTRY	TOTAL	TOTAL		MODEL	INFINIT	OUTFIT	PT-MEASURE	EXACT	MATCH					
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Person	
1	39	40	4.07	1.03	1.08	.41	3.56	1.71	-.14	.12	97.5	97.5	STUDENT_1	
7	39	40	4.07	1.03	.91	.21	.28	-.41	.28	.12	97.5	97.5	STUDENT_7	
29	39	40	4.07	1.03	1.05	.41	1.04	.51	.06	.12	97.5	97.5	STUDENT_29	
31	38	40	3.32	.74	.94	.11	.45	-.31	.29	.16	95.0	95.0	STUDENT_31	
37	37	40	2.87	.62	.99	.11	.86	.11	.21	.20	92.5	92.5	STUDENT_37	
57	37	40	2.87	.62	.99	.11	.86	.11	.21	.20	92.5	92.5	STUDENT_57	
78	36	40	2.54	.55	.80	-.41	.47	-.71	.44	.22	90.0	90.0	STUDENT_78	
13	35	40	2.27	.50	1.00	.11	.96	.21	.24	.25	87.5	87.5	STUDENT_13	
63	34	40	2.04	.46	1.00	.11	.75	-.31	.31	.27	85.0	85.0	STUDENT_63	
84	34	40	2.04	.46	.95	-.11	.85	-.11	.31	.27	85.0	85.0	STUDENT_84	
2	33	40	1.83	.44	1.13	.61	1.05	.31	.18	.28	82.5	82.5	STUDENT_2	
20	33	40	1.83	.44	1.06	.31	.96	.11	.24	.28	82.5	82.5	STUDENT_20	
24	33	40	1.83	.44	.90	-.31	.77	-.31	.39	.28	82.5	82.5	STUDENT_24	
25	33	40	1.83	.44	.82	-.61	.67	-.61	.46	.28	82.5	82.5	STUDENT_25	
40	33	40	1.83	.44	.89	-.31	.78	-.31	.39	.28	82.5	82.5	STUDENT_40	
44	33	40	1.83	.44	1.07	.31	.96	.11	.24	.28	82.5	82.5	STUDENT_44	
60	33	40	1.83	.44	.89	-.31	.78	-.31	.39	.28	82.5	82.5	STUDENT_60	
64	33	40	1.83	.44	1.07	.31	.96	.11	.24	.28	82.5	82.5	STUDENT_64	
76	33	40	1.83	.44	.93	-.21	.67	-.61	.39	.28	82.5	82.5	STUDENT_76	
77	33	40	1.83	.44	1.00	.11	.79	-.31	.32	.28	82.5	82.5	STUDENT_77	
96	33	40	1.83	.44	.81	-.71	.57	-.91	.49	.28	82.5	82.5	STUDENT_96	
43	32	40	1.65	.42	1.11	.51	1.11	.41	.20	.30	77.5	80.2	STUDENT_43	
47	32	40	1.65	.42	.96	-.11	.86	-.21	.34	.30	82.5	80.2	STUDENT_47	
67	32	40	1.65	.42	.96	-.11	.86	-.21	.34	.30	82.5	80.2	STUDENT_67	
85	32	40	1.65	.42	1.03	.21	.88	-.11	.30	.30	77.5	80.2	STUDENT_85	
3	31	40	1.48	.40	.94	-.21	.78	-.41	.39	.31	80.0	78.0	STUDENT_3	
16	31	40	1.48	.40	.81	-.91	.70	-.71	.49	.31	85.0	78.0	STUDENT_16	
74	31	40	1.48	.40	.93	-.31	.72	-.61	.41	.31	75.0	78.0	STUDENT_74	
80	31	40	1.48	.40	.94	-.21	.79	.11	.36	.31	75.0	78.0	STUDENT_80	
92	31	40	1.48	.40	1.02	.21	1.22	.61	.26	.31	80.0	78.0	STUDENT_92	
4	30	40	1.33	.39	1.03	.21	.99	.11	.30	.33	77.5	76.1	STUDENT_4	
14	30	40	1.33	.39	1.05	.31	1.00	.11	.29	.33	77.5	76.1	STUDENT_14	
21	30	40	1.33	.39	1.08	.51	1.15	.51	.24	.33	72.5	76.1	STUDENT_21	
27	30	40	1.33	.39	1.10	.61	.99	.11	.26	.33	72.5	76.1	STUDENT_27	
88	30	40	1.33	.39	.86	-.71	.69	-.81	.48	.33	77.5	76.1	STUDENT_88	
95	30	40	1.33	.39	1.09	.51	1.05	.31	.25	.33	77.5	76.1	STUDENT_95	
33	29	40	1.18	.38	.81	-1.11	.68	-1.01	.53	.34	75.0	74.2	STUDENT_33	
35	29	40	1.18	.38	1.20	1.11	1.12	.41	.18	.34	70.0	74.2	STUDENT_35	
53	29	40	1.18	.38	.81	-1.11	.68	-1.01	.53	.34	75.0	74.2	STUDENT_53	
55	29	40	1.18	.38	1.20	1.11	1.12	.41	.18	.34	70.0	74.2	STUDENT_55	
50	28	40	1.04	.37	.94	-.31	.87	-.31	.41	.35	75.0	72.5	STUDENT_50	
51	28	40	1.04	.37	.99	.01	.93	-.11	.36	.35	70.0	72.5	STUDENT_51	
70	28	40	1.04	.37	.94	-.31	.87	-.31	.41	.35	75.0	72.5	STUDENT_70	
71	28	40	1.04	.37	.99	.01	.93	-.11	.36	.35	70.0	72.5	STUDENT_71	
73	28	40	1.04	.37	.89	-.71	.83	-.51	.45	.35	75.0	72.5	STUDENT_73	
82	28	40	1.04	.37	1.05	.41	1.01	.11	.31	.35	70.0	72.5	STUDENT_82	
26	27	40	.90	.36	1.05	.41	1.54	1.71	.25	.36	77.5	71.2	STUDENT_26	
39	27	40	.90	.36	1.03	.31	.94	-.11	.34	.36	67.5	71.2	STUDENT_39	
59	27	40	.90	.36	1.03	.31	.94	-.11	.34	.36	67.5	71.2	STUDENT_59	
102	27	40	.90	.36	.78	-1.61	.68	-1.21	.57	.36	77.5	71.2	STUDENT_102	
10	26	40	.77	.36	1.00	.01	1.02	.21	.36	.36	75.0	70.1	STUDENT_10	
11	26	40	.77	.36	1.00	.11	.90	-.31	.38	.36	65.0	70.1	STUDENT_11	
19	26	40	.77	.36	.95	-.31	.85	-.51	.42	.36	75.0	70.1	STUDENT_19	
83	26	40	.77	.36	.81	-1.41	.71	-1.11	.55	.36	75.0	70.1	STUDENT_83	
12	25	40	.64	.35	1.06	.51	.98	.01	.33	.37	62.5	69.1	STUDENT_12	

	15	25	40	.64	.35 .95	-.3 .87	-.5 .43	.37	67.5	69.1	STUDENT_15	
	28	25	40	.64	.35 1.13	1.0 1.08	.4 .26	.37	62.5	69.1	STUDENT_28	
	30	25	40	.64	.35 1.27	1.9 1.80	2.7 .07	.37	62.5	69.1	STUDENT_30	
	45	25	40	.64	.35 .83	-1.3 .82	-.7 .51	.37	82.5	69.1	STUDENT_45	
	65	25	40	.64	.35 .83	-1.3 .82	-.7 .51	.37	82.5	69.1	STUDENT_65	
	81	25	40	.64	.35 .85	-1.2 .76	-1.0 .52	.37	72.5	69.1	STUDENT_81	
	86	25	40	.64	.35 1.14	1.1 1.14	.6 .24	.37	62.5	69.1	STUDENT_86	
	97	25	40	.64	.35 1.12	.9 1.04	.2 .28	.37	57.5	69.1	STUDENT_97	
	5	24	40	.52	.35 1.11	.9 1.38	1.6 .25	.38	62.5	68.1	STUDENT_5	
	17	24	40	.52	.35 1.09	.7 1.05	.3 .31	.38	67.5	68.1	STUDENT_17	
	38	24	40	.52	.35 .93	-.5 .87	-.5 .45	.38	67.5	68.1	STUDENT_38	
	46	24	40	.52	.35 .91	-.7 .88	-.5 .46	.38	77.5	68.1	STUDENT_46	
	58	24	40	.52	.35 .93	-.5 .87	-.5 .45	.38	67.5	68.1	STUDENT_58	
	66	24	40	.52	.35 .91	-.7 .88	-.5 .46	.38	77.5	68.1	STUDENT_66	
	79	24	40	.52	.35 1.11	.9 1.14	.7 .27	.38	67.5	68.1	STUDENT_79	
	101	24	40	.52	.35 .82	-1.5 .74	-1.2 .55	.38	77.5	68.1	STUDENT_101	
	6	23	40	.40	.35 1.00	.0 .94	-.2 .40	.39	62.5	67.3	STUDENT_6	
	18	23	40	.40	.35 .85	-1.2 .79	-1.0 .52	.39	72.5	67.3	STUDENT_18	
	22	23	40	.40	.35 .93	-.6 .86	-.6 .46	.39	62.5	67.3	STUDENT_22	
	75	23	40	.40	.35 1.09	.7 1.02	.2 .32	.39	57.5	67.3	STUDENT_75	
	98	23	40	.40	.35 .93	-.6 .86	-.6 .46	.39	67.5	67.3	STUDENT_98	
	99	23	40	.40	.35 .89	-.9 .80	-.9 .50	.39	67.5	67.3	STUDENT_99	
	34	22	40	.28	.35 .94	-.4 1.13	.7 .41	.39	72.5	66.8	STUDENT_34	
	54	22	40	.28	.35 .94	-.4 1.13	.7 .41	.39	72.5	66.8	STUDENT_54	
	49	21	40	.16	.35 .96	-.3 .90	-.5 .45	.40	65.0	66.6	STUDENT_49	
	69	21	40	.16	.35 .96	-.3 .90	-.5 .45	.40	65.0	66.6	STUDENT_69	
	72	21	40	.16	.35 1.03	.3 .99	.0 .37	.40	65.0	66.6	STUDENT_72	
	91	21	40	.16	.35 .99	.0 .96	-.1 .40	.40	65.0	66.6	STUDENT_91	
	32	20	40	.04	.35 1.05	.4 1.04	.3 .36	.40	65.0	66.6	STUDENT_32	
	36	20	40	.04	.35 .88	-1.0 .81	-1.0 .52	.40	70.0	66.6	STUDENT_36	
	52	20	40	.04	.35 1.05	.4 1.04	.3 .36	.40	65.0	66.6	STUDENT_52	
	56	20	40	.04	.35 .88	-1.0 .81	-1.0 .52	.40	70.0	66.6	STUDENT_56	
	8	19	40	-.08	.35 1.03	.3 .98	-.1 .39	.40	60.0	66.6	STUDENT_8	
	89	19	40	-.08	.35 .97	-.2 .90	-.5 .44	.40	65.0	66.6	STUDENT_89	
	94	19	40	-.08	.35 .92	-.6 .85	-.8 .48	.40	65.0	66.6	STUDENT_94	
	9	18	40	-.20	.35 1.29	2.2 1.61	2.8 .09	.41	57.5	67.0	STUDENT_9	
	23	18	40	-.20	.35 1.21	1.7 1.41	2.0 .18	.41	62.5	67.0	STUDENT_23	
	90	18	40	-.20	.35 1.11	.9 1.14	.8 .30	.41	62.5	67.0	STUDENT_90	
	100	18	40	-.20	.35 1.12	1.0 1.25	1.3 .27	.41	62.5	67.0	STUDENT_100	
	48	17	40	-.32	.35 1.20	1.5 1.37	1.8 .20	.41	57.5	67.6	STUDENT_48	
	68	17	40	-.32	.35 1.20	1.5 1.37	1.8 .20	.41	57.5	67.6	STUDENT_68	
	87	17	40	-.32	.35 1.34	2.4 1.69	3.1 .04	.41	62.5	67.6	STUDENT_87	
	93	17	40	-.32	.35 1.12	.9 1.26	1.3 .27	.41	67.5	67.6	STUDENT_93	
	42	14	40	-.71	.36 .96	-.2 .93	-.2 .45	.41	77.5	71.6	STUDENT_42	
	62	14	40	-.71	.36 .96	-.2 .93	-.2 .45	.41	77.5	71.6	STUDENT_62	
	103	14	40	-.71	.36 1.16	1.0 1.36	1.5 .24	.41	67.5	71.6	STUDENT_103	
	41	12	40	-.98	.38 1.21	1.1 1.29	1.1 .22	.41	75.0	75.1	STUDENT_41	
	61	12	40	-.98	.38 1.21	1.1 1.29	1.1 .22	.41	75.0	75.1	STUDENT_61	

	MEAN	26.4	40.0	.94	.41 1.00	.1 .98	.1		73.8	74.1		
	S.D.	6.3	.0	1.00	.12 .12	.8 .35	.9		9.5	7.9		

Figure 2. The Output of Person Statistics

The information that can be obtained from the Measure column is the value of logit person, for person-1 with +4.07 logit shows person with the highest ability (able to do almost all of the questions); person-61 with -0.98 logit indicates person with the lowest ability (lowest in answering the questions correctly). To check the person that do not fit (Outliers or misfit) can be seen in the same Outfit colum with the same criteria when examining the item. From the table above, it can be seen that person-1 misfit because MNSQ> 1.5. this case happens because he was not able to do the question number 5 although it is still relatively categorized as an easy question, while the question with the highest difficulty level (question number 40) can be done correctly. and PT Measure corr <0.4. Beside person-1, person-31 also experienced a misfit because MNSQ <0.5 and PT Measure corr <0.4. This problem occurs because was less careful in work on the questions number 15 and 26. To see the ability of each person, can be seen in the figure below:



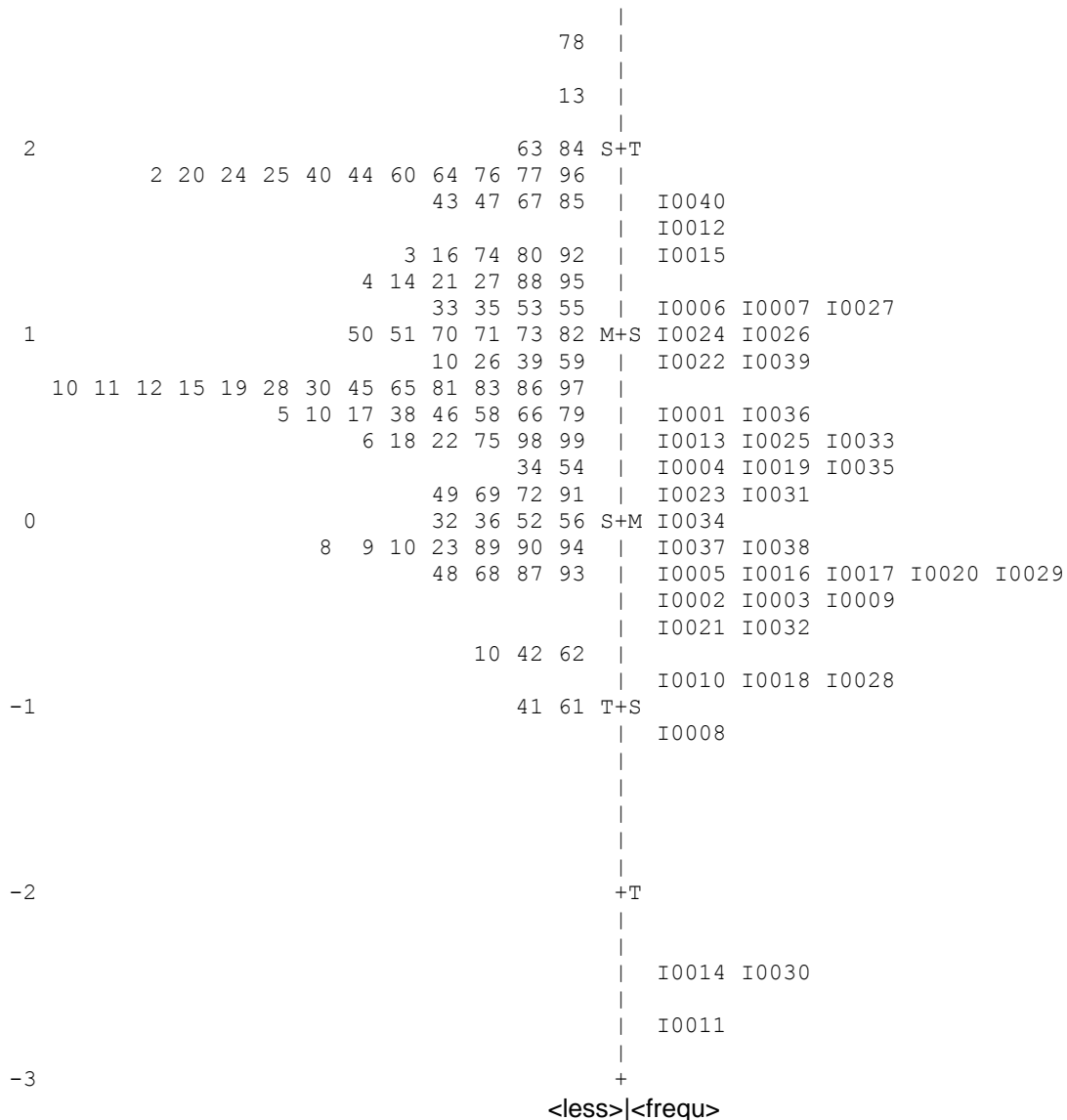


Figure 3. The output of person and item measure

The figure above illustrates the whole person (on the left side) and item (on the right side). Through these figures, we can see the distribution of students' abilities from the highest to the lowest. The distribution of items is also a sequence in accordance with the level of difficulty, looked on the figure that question number 40 is an item with the highest difficulty level and question number 11 is the item with the lowest level of difficulty.

IV. CONCLUSION AND SUGGESTIONS

Based on research discussion above, it can be concluded that the question items of Accounting Final Examination using Winsteps Software show that the Person Reliability is 0,81 while the Item Reliability is 0,93. The value of Alpha Cronbach is 0,82. Person Measure 0.94 logit shows that the average ability of students is above the item, which means that the response patterns of the students is fit. The result also shows that all of the item is fit, so that do not need to be revised or eliminated. Additional information can be obtained are the value of logit person, for the Student-1 with logit +4.07 indicates the person with the highest ability (able to do almost all of the questions); Student-61 with logit -0.98 showed the students with the lowest ability (lowest in answering the questions correctly). Misfit occurs for Students-1 because MNSQ > 1.5 and PT Measure corr < 0.4, this case happens because he could not answer the easy questions correctly, but can answer the difficult item correctly. Additionally, Student-31 also experienced a misfit, where MNSQ < 0.5 and PT Measure corr < 0.4, it happens because he was less careful in

work on the items number 15 and 26. Based on the research results, it can be concluded that most of participants in Accounting Final Examination has the response patterns that fit, so it is assumed that in this Final Examination of Accounting is not found inappropriateness on student response patterns.

V. REFERENCES

- [1] Mardapi, D. (2012). *Pengukuran penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Litera
- [2] Sumintono, B & Widhiarso, W. (2012). *Aplikasi model rasch untuk ilmu-ilmu sosial*. Cimahi: Trimkom Publishing House.

DETECTING STUDENTS LEARNING DIFFICULTIES USING DIAGNOSTIC COGNITIVE TESTS

Yuli Prihatni

Sarjanawiyata Tamansiswa University, Yogyakarta
Email: yuliku7781@gmail.com

Abstract - The study aimed at: (1) Developing diagnostic cognitive test to detect students' learning difficulties, (2) identifying students' learning difficulties, using diagnostic cognitive tests developed. The method used in this research is the development of diagnostic tests. The research was preceded by the preparation of learning continuum, grating test, hierarchy of prerequisite materials and specification tests on materials of science concept were examined through Focus Group Discussion (FGD). Test phase was implemented with test subjects as many as 484 students of class VII SMP in Sleman. Selection of test items to get fit tests using Quest program and M Plus program to test the suitability/fitness model with the attribute matrix Q. The results show: (1) the instruments developed in the form of a multiple choice test with reason, totaling 28 items and seven Q matrix. (2) Cognitive diagnostic test results showed that students in all categories of schools that became the subject of research, experience difficulties on the type of conceptual and procedural knowledge, teachers need to change the nature of students' critical thinking and instill concepts and procedures accordingly.

Keywords: *Learning Difficulties, Tests, Diagnostics.*

I. INTRODUCTION

The diagnostic assessment is essential, to improve the effectiveness of classroom learning. Assessment should produce specific information about the students' cognitive competence. The diagnostic assessment can be performed using diagnostic tests. [1] Test is defined as a number of questions that need answers or statements which must be given a response with the aim of measuring the level of a person's ability or reveal certain aspects of the person subjected to the test". Diagnostic tests can be used to detect students' learning difficulties and weaknesses in accordance with the stages of thinking. Diagnostic test results can be used as a basis for teachers to provide follow up treatment appropriate to difficulties and weaknesses of the student. With diagnostic tests expected destination unknown teaching students can be identified and a known cause. *"Truly diagnostic tests may also help deepen teachers' assessments of their students' conceptual understanding of a specific topic"* [2]. Diagnostic tests can help teachers assess students' understanding, on a particular topic. Cognitive diagnostic test is expected to provide feedback to teachers about students' weaknesses and difficulties in understanding certain competence. Cognitive diagnostic tests performed by the students will be very helpful for teachers, and for researchers investigating cognitive processes. [3] *"A good diagnostic test is one that goes beyond estimations of individuals' overall ability levels. Specific information must be available from student responses that isolates weaknesses or inconsistencies in knowledge and provides a fuller picture of student abilities"*.

The diagnostic test developed in this study is a diagnostic test based on the theory of cognitive Bloom's revised by Anderson and Krthwohl, which Specifies four types of knowledge that is factual knowledge, conceptual, procedural, and metacognitive [4]. Factual knowledge includes the basic elements that should be known to the students to learn the disciplines or to resolve problems in the discipline. Conceptual knowledge includes relationships between elements in a large structure which allows the elements to function together. Procedural knowledge includes how to do something, to practice the research methods and criteria to use skills, algoritme, techniques and methods. Metacognitive knowledge includes knowledge about cognition in general and the awareness and knowledge of cognition itself.

Preparation of cognitive diagnostic tests require precision and accuracy for the purpose of use of diagnostic tests can be achieved. Cognitive diagnostic testing in a psychological or educational setting mainly focuses on at least three aspects of cognitive characteristics:

1. Skill profiles or knowledge lists that are essential in a given cognitive domain.
2. Structured procedural and/or knowledge network.
3. Cognitive processes, components, or capacities.

These three aspects of cognitive characteristics are not exhaustive. Higher-order thinking skills such as cognitive strategy, strategy shifting, and metacognitive skills, should also be included in diagnostic assessment but may be limited by the development of testing techniques at present [5].

II. METHODS

The method used in this research is the development of a diagnostic test through three stages: 1) test planning stage, 2) stage trials, and 3) phase prototype development of computer software. Test phase was implemented with test subjects as many as 484 students of class VII SMP in Sleman. In the first phase, the content validation instrument with FGD. Cognitive diagnostic test instruments that have been arranged in the analysis with a focus group consisting of measurement specialists, cognitive psychologists, experts sain, peers and school teachers who are members of MGMPs Sleman. Analysis of quantitative data using Quest program, CDM and M Plus Program. Quest program is used to obtain items parameter value item difficulty, to complete information on the cognitive diagnostic instruments developed. Diagnostic information obtained with that by linking the item difficulty with learning continuum and cognitive dimensions. Analysis using the Quest program can compare the ability and difficulty that can not be known when analyzed using classical analysis. Thus the results of the analysis Quest program can provide diagnostic information on students' weaknesses and difficulties. Quest program can also be used to determine the fit of items to the model. Thus the Quest program can create a standardized item. CDM program in this study is used to generate the syntax M Plus. M Plus program is used to find more information about the latent class, a way to simulate the size of the matrix Q and the number of test takers. Research Simulation using Monte Carlo simulation study (M Plus). This simulation uses real data responses of participants and attributes through the matrix Q.

III. RESULT

The tests used in this study is an objective test with reason, totaling 28 items. Excess objective tests are as follows: (a) the scope of subject matter that can be asked more spacious compared to other tests. The number of questions asked are generally relatively large. (b) can measure various levels of knowledge. (c) make the assessment score is easy, fast, objective, and may include the scope of materials and extensive material in a test for a class. (d) is ideal to use when participants lot. (e) the reliability of objective tests is relatively higher compared to the test description.

Objective test which include the reason can be to find out whether students master specific competency or not, whether according to his knowledge, whether it is correct steps and reasons for selecting answers. Thus when the student can determine the choice of answers, the students need to know what is the reason. The use of objective tests with reason in general has not been widely used in schools, therefore students were briefed in advance by the teacher before answering the test. Students can choose the answer is to know the reason for choosing that answer.

The instrument has been developed validated by experts through focus group discussions. Overall the results of the validation instrument experts have good and very good value. Data obtained from the FGD then arranged in a Q matrix and Attributes Cognitive Diagnostic Tests in science lessons in junior high. Attribute didefenisiskan as a process or a procedure for resolving students' ability of an item [6]. Attributes are latent characteristics of the test participants/respondents [7]. Instruments cognitive diagnostic tests in science subjects in secondary schools in stacking Q matrix. Each Q matrix consists of 3 and 4 attributes attributes. This study consisted of 7 Q matrix. In this study resulted that the matrix size 3x3 is more ideal than others because it can provide structural information latent class maximum. This is in line with research conducted [8] states that the lowest score on all three information criteria, indicates that the Q matrix was ideal because it can provide information to the maximum class latent

structure. To obtain an attribute profile test participants, M Plus program can provide the attribute profile information in a variety of ways. In this study attribute profile visits using model estimation.

Cognitive diagnostic test results that have been tested in seven junior high schools in Sleman district heat to the material presented in Figure 1.

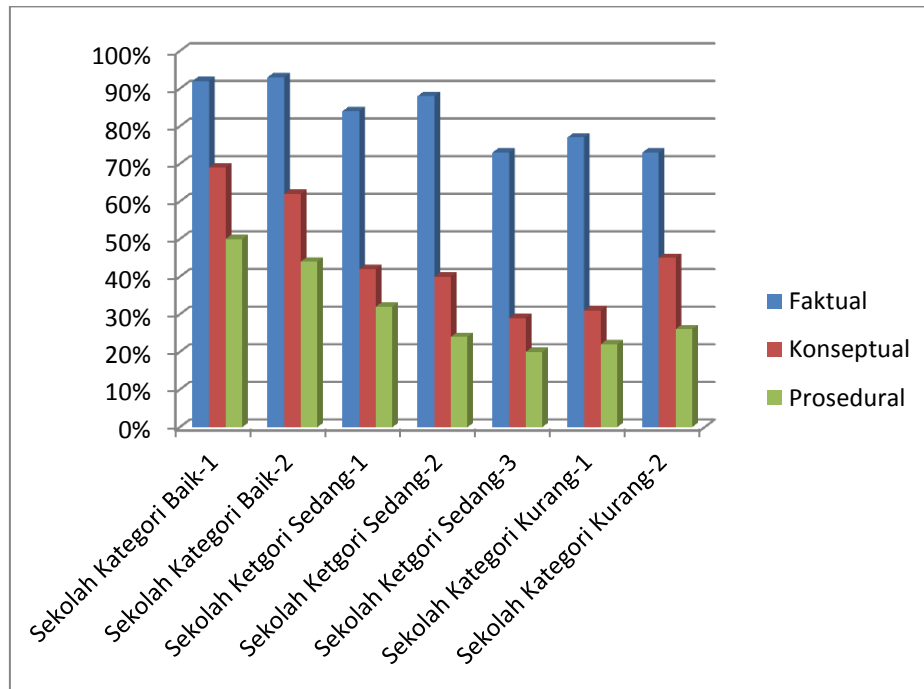


Figure 1. Percentage of Cognitive Diagnostic Test Results

In this study, students in schools with good category, medium category and the category of less, master the test item types of factual knowledge than the type of conceptual and procedural knowledge. And the results of the students' responses showed that procedural knowledge is kind of knowledge that most students are not controlled. This is in line with research [9] states that the student is difficult to connect the conceptual knowledge possessed into procedural knowledge so that students tend to memorize the stage of the matter, is not based on an understanding of the related concepts. Tests to measure the kind of procedural knowledge, usually consisting of cognitive aspects of applications and analysis. It's like a study conducted [10] That almost all students have difficulty in working on those items which included analysis. In science subjects Physics matter of application and analysis should be accustomed given by the teacher so that the students have no problems, but in this study the analysis and application aspects of the students were still very weak. So that teachers in every school needs to enrich and train students by using questions that were able to make students think critically with regard learning continuum in order to improve the quality of learning.

Results of detection weaknesses and difficulties of students in the study visits of this type of knowledge is

A. Factual knowledge

Students experiencing difficulties and weaknesses.

1. Mastering the ability prerequisite heat (temperature and substance)
2. distinguish sublimes and crystallizes.

B. Conceptual knowledge

Students experiencing difficulties and weaknesses.

1. Explaining the concept of heat transfer.
2. Analyzing changes in temperature and states of matter using pictures.
3. Analyzing the application of heat transfer benefits

C. Procedural knowledge

Students experiencing difficulties and weaknesses.

1. Determining the amount of heat in a substance with a certain specific heat.
2. Determining the heat capacity.
3. Measuring the temperature rise on substances that are heated to a certain heat.
4. Analyzing the process of phase transition using charts

Diagnosis difficulties students individually interpreted by cognitive diagnostic test results on line is presented in Figure 2.

3	Menentukan konversi satuan pada besaran kalor	C	Siswa dapat menentukan konversi satuan pada besaran kalor	
4	Menjelaskan sifat-sifat termometrik suatu zat	C	Siswa dapat menjelaskan konsep kalor jenis	
5	Menghitung jumlah kalor pada zat yang diketahui kalor jenisnya	A	Siswa dapat menjelaskan satuan kalor	
6	Menghitung jumlah kalor pada zat yang diketahui kalor jenisnya	A	Siswa dapat menentukan konversi satuan pada besaran kalor	
7	Menghitung kapasitas kalor pada zat yang diketahui massa dan kalor jenisnya	D		Siswa belum dapat menjelaskan satuan kalor dalam SI
8	Menjelaskan pengaruh kalor dalam merubah suhu dan wujud zat	D	Siswa dapat menjelaskan pengaruh kalor dalam merubah suhu dan wujud zat	
9	Menyebutkan perubahan wujud zat yang terjadi dari cair ke padat	B	Siswa dapat menyebutkan perubahan wujud zat yang terjadi dari cair ke padat	
10	Menyebutkan perubahan wujud zat yang terjadi dari gas ke padat	D	Siswa dapat menyebutkan perubahan wujud zat yang terjadi dari gas ke padat	
11	Menyebutkan perubahan wujud zat	B	Siswa dapat menyebutkan	

Figure 2. Test results based Learning students continuum and mastery of the material prerequisites

Thus the interpretation of diagnostic tests can be used by teachers to identify the weaknesses of students and can provide treatment that is in accordance with the needs of students. Factual knowledge with regard to the experiences of students in everyday life. What was experienced by the students will be more easily stored in the brain of the student. Therefore, in the science lesson, the teacher should always relate the subject matter to everyday events.

Teachers should familiarize and train students to work on the problems with this type of procedural knowledge, to acquire procedural knowledge that is connected both with the knowledge conceptual course the students need to be instilled the concepts well, one linking the concept to another concept, and trained how to represent concept with symbols, and trained how to use the rules or procedures to resolve problems of physics.

Diagnostic tests of cognitive development will be more meaningful if it can be used directly by the teacher. In this study, teachers can use the results of diagnostic tests of cognitive as a base to provide follow up treatment appropriate to the learning difficulties of the students. By using computer software that comes with a password, for teachers to get information: (1) the number of items answered correctly and items that can be answered students based on the type of knowledge, (2) information about the material that has been mastered and not mastered students based learning continuum and material prerequisites.

This research can be answered, why the students have not mastered certain competencies, and can interpret the results of diagnostic tests of cognitive science subjects at junior high, so as to diagnose students' learning difficulties. Thus, the difficulty of students in science learning

materials physics of heat will be solved, because the latent class known any material that is not controlled by the student. Teachers can provide treatment mainly lies the weakness and learning difficulties experienced by students and the possibility of achievement that can be developed.

Cognitive diagnostic test developed in this study combined computer software can help teachers to give lessons in accordance with the thinking of students thus increasing the quality of science teaching. Diagnostic tests developed have the ability to test the student's ability to answer the questions correctly to evaluate the automatic response, then generate a list of weakness or difficulty found. The feedback provided can be used by teachers to improve student skills by focusing on reported weaknesses, and shows suggestions for improvements to enhance the learning process.

Cognitive Diagnostic test results revealed that, in all categories of schools that is the subject of the study had difficulty, on a test type of conceptual and procedural knowledge. Teachers need to change the nature of students' critical thinking and instill concepts and procedures accordingly.

IV. CONCLUSION

Cognitive diagnostic test developed in this study is an objective test with reasons, totaling 28 items. This cognitive diagnostic tests can be directly used by teachers to produce information that can diagnose student difficulties, so that teachers can provide follow-up and appropriate treatment in order to improve the quality of learning. The results showed that in every school, students have difficulty at the time of taking the test with the type of conceptual and procedural. Teachers need to improve learning by training students in completing the test with the type of conceptual and procedural knowledge.

REFERENCES

- [1] Djemari Mardapi. (2008). Teknik Penyusunan Instrumen Tes dan Non Tes. Yogyakarta: Mitra Cendikia Offset.
- [2] Russell, M., & O'dwyer. L.M. (2009). Diagnosing students' misconceptions in algebra: results from an experimental pilot. *Behavior Research Methods*. 41(2),414-424.
- [3] Gorin, J.S. (2007). Test construction and diagnostic testing. Dalam Leighton, J.P., & Gierl, M.J.(eds). *Cognitive diagnostic assessment for education theory and applications*.(pp 174). Cambridge. Cambridge University Press.
- [4] Anderson, L. W., & Krathwohl, D.R. (Eds). (2001). A taxonomy for learning, teaching and assessing: A revision of Bloom's *Taxonomy of educational objectives*: Complete edition, New York: Longman.
- [5] Xiangdong, Y., & Embretson, S.E., (2007). Construct Validity and Cognitive Diagnostic Assessment Dalam Leighton, J.P., & Gierl, M.J.(eds). *Cognitive diagnostic assessment for education theory and applications*.(pp 119). Cambridge. Cambridge University Press.
- [6] Gierl, M.J. (2007). *Making Diagnostic Inferences About Cognitive Attributes Using the Rule Space Model and Attribute Hierarchy Method*. *Journal of Educational Measurement*, 44(4), p.327.
- [7] Rupp, A.A., Templin, J. & Henson, R.A. (2010). *Diagnostic measurement: Theory, methods and applications*. New York: The Guilford Press.
- [8] Kusaeri. (2012). Pengembangan tes diagnostik dengan menggunakan model DINA untuk mendapatkan informasi salah konsepsi dalam aljabar. *Disertasi doktor*, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- [9] Oksuz, C. (2007). Childrens understanding of equality and the equal symbol. Adnan Menderes. Diambil dari cumalioksuz@adu.edu.tr
- [10] Fauzan. (2010). Pengembangan tes diagnostik kesulitan belajar matematika di SMA. *Disertasi doktor*, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

Development and Implementation of Higher Order Thinking Skills Instruments in Physics Education

A.Halim^{1*} and Yusrizal²

^{1,2} Department of Physics Education, Teacher Training and Education Faculty,
Syiah Kuala University, Aceh, Indonesia.

*e-mail: bdlhalim@yahoo.com

Abstract—The purpose of research is to develop a test that can be used to assess Higher Order Thinking Skills (HOTS) on high school students. Development of test adopted the ADDIE's model and the results of the development an essay test is called a Higher-Order Thinking Skills (KBTT) instrument on Static Fluid material. The results of the pilot test the KBTT limited to 59 students showed that (i) the average index of difficulty is 0.65 with medium category, (ii) the average index of discrimination is 0.36 with enough categories, (iii) the average value of reliability is 0.91 with the category of very high, and (v) the validity of an average of 0.70 with a high category. While the results of the implementation of the KBTT on 54 students showed that the indicator of the KBTT have high contribution (0.65) to the total score of the KBTT and also has a high correlation (0.83) to the Understanding of Concepts (PK). This indicates that the instrument of the KBTT have a good ability to measure the cognitive domains of C1 through C6. The implication of this result is expected to teachers and lecturers in order to be able to use higher order thinking skills instruments to measure the achievement of physics learning

Keywords: *physics, learning, assesment, skills, higher, instrument, thinking.*

I. INTRODUCTION

Efforts to improve the quality of education in this country remains a major priorities, because the quality of education the people of Indonesia compared with other countries is still very low. Anis Baswedan, menteri pendidikan dan kebudayaan, mengatakan kondisi pendidikan di Indonesia berada pada tahap yang sangat kritis, ini sesuai hasil survei yang dilakukan oleh PISA (Programme for International Study Assessment), dimana kualitas pendidikan pada kedudukan 64 dari 65 negara yang disurvei. Whereas, human resources if based on the Human Development Index (HDI), UNDP survey results show the people of Indonesia is 108 out of 187 countries surveyed, while Malaysia 62 and Thailand on 86 [1]. This fact shows that the quality of the nation can be increased through improvement of Human Resources (HR). The formal education from elementary school to the PT is the agency to produce human resources that have the ability to think low level or *Lower Order Thinking* (LOT) and high-level thinking skills or *Higher Order Thinking* (HOT).

Higher order thinking skills (HOTS) has been defined in different forms. King said that the Higher order thinking skills include the ability to think creatively, metacognitive, reflective, logical, and critical [2]. Higher order thinking is based on the version level thinking skills such as distinguishing, application and analysis simple and cognitive strategies associated with knowledge ago. Foster as a director of ACER, said that the Higher order thinking skills including literacy, critical, critical numeracy (ability in mathematics), and cross-cultural competence [3]. Whereas Wheeler & Haertel, concluded that the higher order thinking skills are grouped into two different contexts [4]. The first context, the thought process needed to solve problems and make decisions on every issue in everyday life. The second context, the mental processes required for success in teaching, including activities to compare, evaluate, reason (justification), and make inferences. Brookhart distinguish the definitions of Higher order thinking in three categories, (1) defines a Higher order thinking from the point of view of the term displacement (transfer), (2) define it from the perspective of critical thinking, and (3) define it in the context of problem solving [5]. According to Costa higher order thinking skills can be divided into four types, namely critical

thinking, creative thinking, problem solving, and make a decision [6]. Based on the above definition can be said to be some higher order thinking skills is an activity involves thinking and mental processes at the highest level, including processes, analysis, comparing, inferential, evaluation, and creativity in order to solve problems in everyday life.

Many forms of tests have been used to measure higher order thinking skills. Form of higher order thinking skills tests as a form of learning products including; multiple choice, essay exams, and a portfolio that is often used for this. Among the three forms that have a higher validity and reliability as well as have a significant impact on learning is a portfolio assessment. Besides, higher order thinking skills can also be viewed as a learning process. The measurement is focused on the procedural implementation capabilities at the time of learning activities using observation sheet and questionnaire [7]. Test essay or answer descriptions have been developed to measure critical thinking (part of a high level of skill) with the validity of an average of 0.60 (medium), reliability 0.55 (medium), index of different power averaging at 0.35 (enough), and the level of difficulty of average -rata 0.45 (medium) [8]. Observation sheet also used to measure critical thinking skills at the time a student learning activities with guided inquiry method [9].

Based on the views of experts and relevant research results above, will be developed through this research essay test instrument to measure higher order thinking skills. Having tested the validity and reliability by experts, pilot test carried out on 59 students of class XI Madrasah Aliyah (MA) in the static fluid material.

II. METHOD OF RESEARCH

This study used a technique for Research and Development (R & D) by taking the ADDIE model of [10]. ADDIE Model has five phases, namely Analysis, Design, Development, Implementation, and Evaluation, the fifth phase is shown in Figure 1 below. In the context of this study, ADDIE model application includes;

Analysis stage is an activity to obtain field information that does not fit the purposes of the application of PBL to increase the high level of skill with the target reached. Information obtained through the analysis phase that targets school teachers in this study have not been using tests to measure higher order thinking skills. Whereas 2013 curriculum requires teachers to use learning model that can increase higher order thinking skills and how to measure it.



Figure 1. ADDIE's model
(souce: Dick & Carey, 2014).

Design stage is planning activities to develop measurement instruments higher order thinking skills in a static fluid materials at schools targeted research. The instrument was developed tailored to the needs of teachers in targeted schools.

Develop stage is the stage of development of higher order thinking skills instrument in a static fluid material. Stages of development include;

- Redefine higher order thinking skills in real and operational.
- Identify indicators of higher order thinking skills
- The validity of the expert to the indicator obtained
- Prepare guideline of multiple-choice test on the static fluid material.
- Writing test items (MCQs).
- The validity of the experts of the items were written
- The test instrument on 59 students of class XI Madrasah Aliyah (MA) in the static fluid material.
- The result of the development of measurement instruments formed Higher-Order Thinking Skills (PKBTT).

Implementation stage is the stage of the use of instruments PKBTT conducted by researchers at 26 class XI experimental group and 27 students of class XI control group at Madrasah Aliyah Rukoh for static fluid material. In the experimental group applied learning models Problem Solving (PS). Design implementation of the instruments shown in the following tabel.1.

Table 1. Design implementation of the instruments PKBTT

Group	Pre-test	Treatment Method	Pos-test
Experiment	O1	PS	O2
Control	O1	Conventional	O2

Evaluation stage is the stage of assessing the effectiveness or usefulness level essay test instrument to measure higher order thinking skills in students Madrasah Aliyah Rukoh Banda Aceh. The first evaluation techniques, comparison of measurement results before treatment and after treatment learning model. Second, calculate the correlation between the test results of PBKTT and test results of Concepts Understanding (PK) in the Static Fluid.

III. RESULTS AND DISCUSSION

The results of pilot tests show that (i) the index of difficulty average of 0.65 with medium category, (ii) index of discrimination average of 0.36 with enough categories, (iii) the reliability value average 0.91 with a very high category, and (v) the validity of an average of 0.67 with a high category. These values are compared with the results of research Guspitasari not much different, with the validity of an average of 0.60 (medium), reliability 0:55 (medium), index of different power averaging at 0.35 (enough), and the level of difficulty of the average 0:45 (medium) [7].

Based on Bloom's taxonomy revised edition, which included cognitive indicators of higher order thinking is analytical (C4), evaluation (C5), and creative (C6) [11]. The results of the pilot test showed that the indicator KBTT test "analysis" took the biggest increase in sample experimental group (0.85), and indicator "creative" an increase in the lowest (0.82). These analyzes showed that the indicator has the best contribution to support the overall ability KBTT instrument. This difference is due to the material content and the methods used more students are actively involved in the implementation of learning. This is in accordance with the opinion of Eko, who says that the learning model Problem Solving (PS) to enhance understanding of concepts and higher order thinking skills (KBTT) [12]. Besides, Guspitasari and Sodjatmoko also said that the process of learning that involves students actively fosters thinking skills on students [7,13].

These indicators have varying contribution to the instrument of KBTT, as shown in Figure 2 below. Among the three indicators of KBTT instruments, the highest contribution is the analysis (0.72), then kreartif (0.59), and the final evaluation (0.49). This indicates that the instrument KBTT developed in this study is more predominant activity "analysis" conducted by the person taking the test, compared with the evaluation or creative activity.

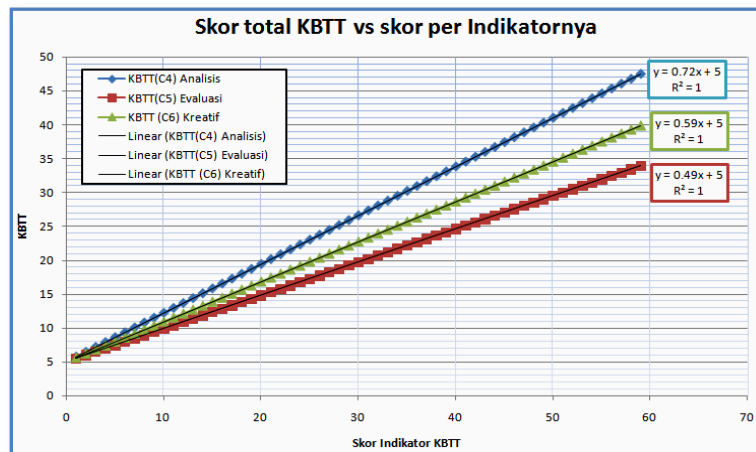


Figure 2. Correlation between the score of KBTT and its indicators

While the implementation of the instruments PK TT conducted on 26 students of class XI-IA3 as an experimental group and 27 students of class XI-IA2 as the control group.

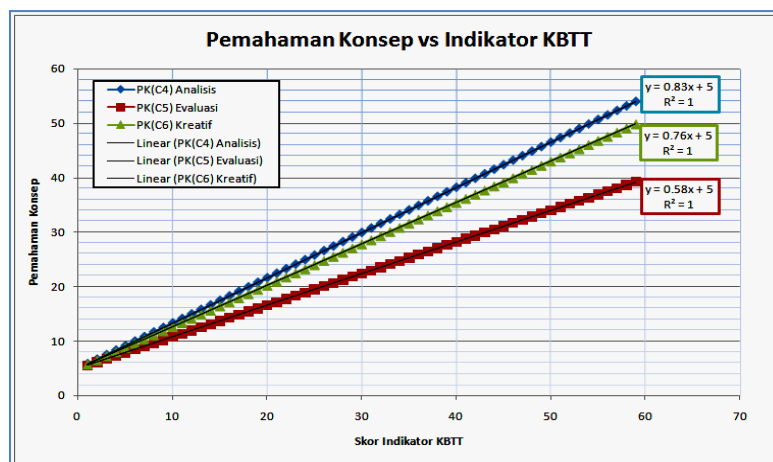


Figure 3. Correlation between PK with the indicators of KBTT

The implementation results show that the instrument can measure PKBTT relationship (correlation) between the higher order thinking skills (KBTT) with understanding of concepts (PK) Fluid Static in middle category (0.33). More detailed understanding of concepts (PK) also has a strong correlation with each indicator KBTT as shown in Figure 3 above. Low ability PKBTT instrument to reveal the relationship between KBTT with PK caused by the all content or the content of each test is a bit different

IV. IV CONCLUSION

The study had two main goals, namely the development of measurement instruments Higher-Order Thinking Skills (PKBTT) and the level of reliability (effectiveness) PKBTT instrument in measuring the higher order thinking skill improvement. The results showed that the indicator KBTT for "analysis" took the biggest improvement, and indicator KBTT for "creative" an increase in the lowest. The implementation results show that PKBTT instrument can measure the link between higher order thinking skills (KBTT) with a mastery of concepts (PK) Fluid Static by students in the category.

ACKNOWLEDGMENT

Further thanks go to everyone who has supported the implementation of this study. Especially to our students (Sabaruddin) who helped carry out the research properly and sincerely, and also

to the manager and administrator of the Master of Science Education which had contributed to the successful implementation of this idea.

REFERENCES

- [1] Muchsini, B (2016). Integration of Higher Order Thinking Skills in Assessment Instrument Accounting Computer at Higher Education *PROSIDING ICTTE FKIP UNS 2015*, Vol 1, Nomor 1, Januari 2016
- [2] King, F.J., Ludwika, Ph.D., Goodson, M.S., Faranak Rohani, Ph.D. (2014). Higher Order Thinking Skills. A publication of the Educational Services Program, known as the Center for Advancement of Learning and Assessment diakses pada www.cala.fsu.edu. Januari 2016.
- [3] Foster, M., (2004). Higher Order Thinking Skill. *Research Developments*, Vol. 11 Art. 1 diakses pada <http://research.acer.edu.au/resdev/vol11/iss11/1> Januari 2016.
- [4] Wheeler, P. and Haertel, G. (1993). *Resource Handbook on Performance Assessment and Measurement*. Berkeley, CA: The Owl Press.
- [5] Brookhart, S. (2010), *How to Assess Higher Order Thinking Skills in Your Classroom*, ASCD, <http://www.ascd.org/Publications/Books/Overview/How-to-Assess-Higher-Order-Thinking-Skills-in-Your-Classroom.aspx>
- [6] Costa, A.L. (Ed.). (1985). *Developing Minds: A Resource Book for Teaching Thinking*. Alexandria, VA: Association for Supervision and Curriculum Development.
- [7] White, E.M (1991) Assessing Higher Order Thinking and Communication Skill U.S. In College Graduates Through Writing, Presented to the *National Assessment Conference Washington*, DC: November 17-19, 1991, California State University, San Bernardino.
- [8] Gupitasari, Gita (2015) Penurunan Kecemasan dan Peningkatan Kemampuan Berpikir Kritis Matematis Siswa SMP Melalui Model Pembelajaran Knisley. S2 thesis, Universitas Pendidikan Indonesia.
- [9] Sochibin, A, Dwijananti, P., & Marwoto, P. (2009). Penerapan Model Pembelajaran Inkuiri Terpimpin untuk Peningkatan Pemahaman dan Ketrampilan Berpikir Kritis Siswa SD. *Jurnal Pendidikan Fisika Indonesia* 5 (2009): 96-101
- [10] Dick, W., and Carey, L. (2014). *The Systematic Design of Instruction*. N.Y. Pearson Education, 8th ed
- [11] Anderson, L.W., & Krathwohl, D.R. 2001. A taxonomy for learning teaching, and assessing: a revision of bloom's taxonomy of educational objectives (Rev. ed). New York: Addison Wesley.
- [12] Eko, S. W. 2012. Penerapan Model Pembelajaran Problem Solving Fisika Pada Pembelajaran Topik Optika Pada Mahasiswa Pendidikan Fisika. *Jurnal Exacta*. X(2):111-114.
- [13] Soedjatmoko. 1991. Mencari Strategi Pengembangan Pendidikan Nasional Menjelang Abad XXI. Jakarta : Grasindo.

DEVELOPING PICTURE SERIES AND VOCABULARY TO INCREASE ENGLISH SPEAKING SKILL

Agustina Ellyana, Ketut Martini and Agus Risna Sari
Master of Educational Technology, University of Lampung
Jl. Sumantri Brojonegoro Gedung Meneng Bandar Lampung
Email: agustinaellyana@gmail.com

Abstract-The objective of this research is to analyze theoretically about the effectiveness, efficiency and attractiveness of the picture series for learning vocabulary mastery in English speaking skill. Based on the theory and research findings from various research journals, vocabulary is the center of expertise in language skills. Therefore, learning it is something very important. This study is a research and development that is used to produce certain products that have been tested effectiveness. The researcher uses this method because the resulting product is the development of pictures that already exist and are often used in the process English speaking skill.

The using of visual media or picture series in the learning process, will increase the motivation to speak English. Picture series that is used can serve to deliver messages more effective, efficient, and attractive. The message delivered is created into symbols of visual communication. The analysis showed an increase in vocabulary mastering through developing picture series is effective in improving English speaking skills. The learning process will be more active, and learners' motivation can be increased by factual learning experience through picture series.

Based on this research, it can be concluded that the using picture series can improve English speaking skill of students significantly.

Keywords: picture series, vocabulary effective, efficient and English speaking skill.

I. INTRODUCTION

Since English is taught as a compulsory subject in Indonesia, the students are expected to learn it well and the teacher is required to be able to teach it well too. It functions as a means for developing student's knowledge in science, technology, arts and culture. The functions are aimed at enabling the students to be creative, skillfull, and become good Indonesian citizens to participate in the country development (Kurikulum SMP, Depdiknas, 2006).

To come to the above functions, the government has set an institutional goal that the students have certain proficiency level of skill and knowledge of English at the end of SMP schooling. The language skills concerned are listening, speaking, reading and writing. However, the writer has found that many students have lack of vocabulary so they cannot speak in English well.

Speaking means to express thoughts verbally and make other people talk to understand what he had in mind (Djiwandono 2008: 118). Speaking is a natural and integral things that we forget how it was first obtained and able to do it so when we want to master a foreign language we must study it from the beginning. In speaking English, students do not have a good ability to convey the contents of the message to others. It is aligning its capabilities right between what you have in mind with what he said, so that someone who listen it may have the same meaning with the speaker's intention. Generally, students realize that the speaking ability is a means to communicate, or provisions to continue their studies to a higher level. But most students often have difficulty with their English speaking tasks.

The difficulty may be the selection of appropriate vocabulary, less fluent, as well as less obvious in expressing their ideas. Vocabulary as one element of language plays an important

role in speaking activities. Through words, they can expressed his thoughts, ideas and feelings clearly.

The low skills of English speaking students often complained not only by primary but also by the junior high school teachers. The construction and habituation of using English daily well needed to be implemented continuously since the elementary and junior high schools for the 9-year basic education. On the other hand some teachers realizes that they have some weakness in their performance, especially in leading students to have high motivation to speak in English. The selection of instructional media can not become more active, less optimal learning management, lack of opportunity given to the students to practice their capability in English speaking. Data student achievement in the school year 2014/2015 show that the speaking skills still low with an average of 62 (KKM 70). This level is still below compared with the other skills that are listening 72.5, reading 77.5 and writing 73. Based on a review of the achievement of the score of students 'speaking skills, it is known that the low score of students' speaking skills are closely related to the low mastery of the aspects of their linguistics and sociolinguistics. Students still have difficulties in developing their idea of the themes discussed in speaking skills. The limited vocabulary is also made difficulty to the students to express their ideas. In addition, students who lack understanding about the various functions of expression of speech and linguistic context are also difficult to increase their speaking skills.

Based on the results of preliminary studies conducted by the writer, during the last 2 years that there were many students who have lack of the vocabulary and so they have low ability in English speaking, especially students of class VIII SMPN 2 Kalianda.

From the data above, it can be concluded that the vocabulary and the ability to speak English class VIII SMPN2 Kalianda South Lampung is still low. When examined in more depth, a factor of the student as the dominant factor in learning to speak. Tthe writer suspect the factors considered as causes of low speaks any skill is the lack of knowledge about the applicable rules of language, lack of vocabulary and limited knowledge or experience (schemata) that will be delivered to the listener or listeners. The quality of a person's language skills obviously depends on the quantity and quality of its vocabulary.

According to Piaget (cited by Roebiyarto, 2010: 1) the human mind has a structure called a schema or schemata (plural) are often called cognitive structures. By using the schemata of a person to adapt and coordinate its environment to form new schemata, namely through the process of assimilation and accommodation. Schemata are formed through a process of assimilation and accommodation that is called knowledge. Assimilation is a cognitive process in which a person integrates information (perceptions and concepts) or a new experience into cognitive structure (schema) that is already owned by someone.

One of the most important insights of the schemata theory is that meaningful learning requires the active involvement of learners, which have very much experience and prior knowledge for use in understanding and uniting new information. What is learned from every experience depends in large part to the schemata that have been applied to the experience. (Slavin 2006: 250) The using of the schemata in the language learning process must be based on the communicative competence that has been owned by someone. Communicative competence is an aspect that enables inter-personal messages can be delivered and interpreted. One of the most important insights of the schemata theory is that meaningful learning requires the active involvement of students, which has so much experience and knowledge prior to use in understanding the new information. What is learned from every experience depends largely on the schemata that have been previously owned (Slavin 2006: 250). Speaking skills students need skills to compose words and sentences either from treasury, the nature and purpose of speaking, the accuracy of speech, intonation, gesture, targeting accuracy and precision of selecting the words used. Students who have knowledge of vocabulary includes:

a common word, special, denotation, connotation, synonyms, antonyms and schemata as well as they would have the ability to express good ideas in speaking. If it cannot be function in certain, the students will have difficulty in expressing ideas / ideas. Students will be difficult to speak English they lack of vocabulary and schema. It has been suggested that knowledge of vocabulary and schema relevant with the level of students' speaking ability.

The better level mastering of vocabulary knowledge and schemata may make students improve their speaking ability. During this time, the teaching of English in speaking skills in SMPN 2 Kalianda only use textbooks and worksheets that are available in the school library.

Both of these types of books already out of date, not in accordance with the needs of students who want the latest materials in language learning. The instructional media only use images contained in both the instructional materials. It causes learning feels monotonous, not interesting for the students, because this situation cannot raises the students' creativity it is not factual. Referring to some of the assumptions above, the writer interested in conducting research whether the schemata, vocabulary and instructional media using a series of picture can improve students' speaking skills. Therefore, this study departed from the assumption that the acquisition of schemata influence on conversational skills. Both are thought to have a very close relationship. Besides vocabulary, human also considered influential on speaking skills so that between schemata, vocabulary and speaking skills are interconnected and influenced. Prior knowledge students are generally called schemata, is expected to expand and enrich the vocabulary after being given a series of picture, so that they also expected to increase their speaking in English.

II. RESEARCH METHODOLOGY

This study is a research and development (Research and Development) is a method used to produce certain products that have been tested effectiveness (Sugiyono, 2013: 407). The writer uses this research method because the resulting product is the development of picture as a media that already exist and are often used in the process of learning to speak English.

In the design of Research and Development proposed by Borg and Gall (1996: 715-716) includes 10 stages, while the small-scale research can be done through the more simple stages. Furthermore, researchers can stop the research on the stages as needed, because the stages are implemented fully, is relatively expensive, very wide scope, and spend a long time (Sukmadinata, 2005: 169). Referring to the above literature, of the ten steps developed by Borg and Gall, researchers simplify and combine some steps so there is only five-step research. This is done because of the limitations, both in terms of time and cost. The next five steps combined with the design concept of learning from Dick and Carey (2001). The five steps being taken are: (1) a preliminary study; (2) planning of product development; (3) development of draft product; (4) The revision of the product, and (5) the dissemination and implementation.

III. FINDING AND RESULT

Due to the research about the advantages of pictures series as instructional media to increase students' English speaking, here are some of the results of relevant research that has been done:

1. Santi (2012) game card number display can be used to improve the cognitive abilities of children. For kindergarten teacher in the learning of cognitive abilities is advisable to use interesting pictures so that children do not feel bored, children interested in learning and provide opportunities for children to explore the potential and intelligence of children. 2. Ubaidiyah (2011) game count with a media card numbers can improve cognitive abilities of children in group A in TK Dharma Wanita Persatuan II Parasrejo Pasuruan, is evident from the results obtained child can be seen from the average results of observations of children ranging from pre-action (49, 8) with a percentage (30%), increased first cycle (65.05) with a percentage (65%) and increased again the second cycle (81.6) with a percentage (90%) which is constantly increasing. 3. Darsono (2007) the use of learning media in the form of still images (visual) in learning can improve the quality of learning organized teachers to be more concrete and meaningful (meaningful learning), which is characterized by the involvement of students in active, creative, effective and fun and can improve acquisition of student learning outcomes.

4. Daroah (2013) achieved the language skills of the students group B1 RA. Perwanida 02 Slawi further increased in comparison with the previous in which the language development of children is only 50%, but after the practice of classroom action research through storytelling using audio-visual media, in the first cycle increased to 75%, and in the second cycle increased reached 85%. Based on the results of this study indicate that: learning through storytelling with audio-visual media can be said to be successful in order to improve the language skills of children. 5. Yanti (2015), the student's skills in storytelling becomes more increased because students easier in telling a story with the help of pictures. By using media images storytelling series on learning, student learning outcomes has increased significantly. Improving student learning outcomes from the first cycle to the second cycle also experienced an increase in the amount of 63.5%.

IV. CONCLUSION

Based on the description in the discussion we concluded that learning to speak English using the pictures serie can improve students' vocabulary and schema. By using the pictures serie on English language learning, student learning outcomes in speaking skills has increased significantly.

REFERENCES

- [1] Daroah. 2013. Improving Language Ability Through Storytelling Method *the Audio Visual Media in Group B1 RA Perwanida 02 Slawi*. Semarang: Semarang State University.
- [2] Darsono. 2007. *Journal of Education LPMP*. Bandar Lampung: LPMP.
- [3] Dick, W. Carey, L. Carey, J. 2005. *The Systematic Design of Instruction* (6th ed.). Boston: Pearson.
- [4] Gall, D. Meredith. Borg, R. Gall, P. 1996. *Educational Research, an Introduction*. Sixth edition. New York: Longman.
- [5] Government Regulation No. 19 Year 2005 on the National Education Standards. Jakarta.
- [6] Santi, Rita. 2012. *Application Builder To Score Card Games Improve*
- [7] *Children's Cognitive Ability Group A TK PGRI 4*. Kolursari-Bangil.
- [8] Sardiman, A. M. 2009. *Interaction And Motivation and Learning*. Jakarta: PT RajaGrafindo Persada.
- [9] Slavin, Robert E. 2000. *The cooperative learning Theory, Research, and Practice*. Second Edition. Noedham height: A. Simon and scuster Company.
- [10] Sugiyono. 2013. *Quantitative Research Methods, Qualitative and R & D*. Bandung: Publisher Alfabeta. Sukmadinata, N. 2005. *Educational Research Methods*. Bandung: Teens Rosdakarya. Ubaidiyah, Liza. 2011. *Application of the game count with a media card numbers to improve the cognitive abilities of children in TK Dharma Wanita Persatuan II Parasrejo*. Pasuruan.
- [11] Yanti, Nofvi. 2015. *The increase in the Child Language Development Through Use of the Media Image In Raudhatul Athfal Al Hikmah Bandar Lampung*. Thesis. Postgraduate Education Technology. Bandar Lampung: Lampung University.

Indonesian Adaptation Scale of Zung Self-Rating Anxiety Scale (SAS)

Alfiannor Luthfi Hasain

Faculty of psychology

Syarif Hidayatullah State Islamic University Jakarta

fianhasain@gmail.com

Abstract-This article aims to report the students' anxiety scale when they perform at school using a scale adapted from Zung Self Rating Anxiety Scale. In adaptation proses of Zung self-rating anxiety scale, the 18 items was translated twice and then back translated. The final form of items was distributed to 154 students of SMP YAPAN, Depok, West Java. Construct validity test on this instrument uses Confirmatory Factor Analysis (CFA) method. The result of construct validity test on students' anxiety instrument when they perform at school shows that all items are unidimensional and can be accepted except 3 items, which are item 11,15, and16. Practically, this scale is beneficial to know the students' anxiety level during school performance so that all teachers can decide on proper treatment for them.

Keyword: *Adaptation scale, SAS, Confirmatory Factor Analysis*

I. INTRODUCTION

This article focuses on the discussion about anxiety adaptation scale. The case investigated by the author is the effect of new environmental situation on somebody that makes him or her feel uncomfortable such as when a student is instructed by his instructor to come forward to perform singing or explain materials from the lesson. Such condition will negatively influence the students' performance; it is seconded by a research conducted by Cassady and Jhonson (2001) who said that if the anxiety level is high, it will negatively influence the performance. Eysenck, Derakshan, Santos, and Calvo (2007) also said that anxiety decreases control over the goals.

Researchers about anxiety has been done a lot, even Allen, Lerner, and Hinrichsen(1972) have suggested that by looking at anxiety test result, we can see the opportunity of potential academic result. Although a topic about anxiety have been researched a lot up until recently, the anxiety variable is still becoming the negative effect of someone's performance because the evidences found related to the anxiety factors are still general, not specified to certain condition, thus the intervention is still on general level. Concerning this condition, the author finds it interesting to understand deeper about anxiety variable by adapting Zung Self-Rating Anxiety Scale (SAS) into anxiety context in which students perform at school. The focus of this adaptation result is expected to be able to be used to understand the anxiety level of Indonesian students especially junior high school students and give some suggestions regarding the proper treatment for the students who experience it.

The adaptation of Zung Self-Rating Anxiety Scale (SAS) is based on theory of performance anxiety as a terrifying condition about non-pleasant things in which fear appears as a reaction of the real danger (Provost in Leao, 1998). Kenny et. al. (2004) explains performance anxietyas a kind of disorder that affects individual in managing his work, such as exam, competition, and public performance. It happens in many fields including academic achievement, public speaking, sport, and art such as acting, dance, and music. From the two conceptual definition, the author perceives that performance anxietyis a kind of disorder manifested in a feeling of uncertainty such as fear of the consequence.

Provost (in Leao, 1998) explains about the basic nature of Performance Anxiety, that is the kind of anxiety that appears from the interaction between three psychological aspects: cognitive, behavior, and physiology. Cognitive components include minds and mental image relationship about danger. Components of behavior cover the tendency to avoid or hide oneself from 'danger' which is involved in the technical problems of performance, such interpersonal problem with teachers, friends, and parents and problem to deal with difficult matters. Physiological components include body reaction to the danger stimulation that appears when there is fear of what will happen, lack of control, inability to think clearly, and prediction of failure.

From the above theoretical consideration, the writer concludes that there are three basic aspects in the performance anxiety that will be measured and collaborated with Zung Self-Rating Anxiety Scale (SAS):

- A. Cognition, where a student experience loss of concentration, confusion, and failure to remember what he should do during performance.
- B. Physiology, in which a student experiences breath problems, digestion problem, increase of heart rate, high muscle tension, sweating, and dry mouth during performance.
- C. Behavior, in which a student experience tremor, body constantly shaking, unorganized movement, and failure to perform.

II. METHOD

Adaptation process was conducted using five steps according to Beaton et al., (2000). The following is the process:

- A. The author translates the scale into Indonesian. The translation process is done by two people. The first has earned master degree in English education. The second is a bachelor in psychology. This is done to obtain comprehensive translation from the combining result of a linguist and someone who understand the concept of the scale.
- B. After receiving the translation from first translator (T1) and the second translator (T2), the author then synthesize P1 and P2 results. If there is gap between the two translation, the author will choose the one that has the most suitable meaning with the scale. In the process of selecting the translation result, the author also considers cultural factors, scale target, and theoretical foundation that has been established.
- C. The author retranslates the text to the original language of the scale. This is done to see if there is any difference in meaning if the scale is Indonesian. If there is difference in meaning, the author should review the item.
- D. After fixing the translation by considering the result of back translation, the author discusses the result with some fellow author in the Magister of Psychological Science of UIN Jakarta who understand about the scale concept.
- E. The scale that had been agreed upon through discussion is then tried to some respondents. This phase is done to know if the scale being adapted can be understood by the respondents.

The data distribution was done by coming to the field. The author distributed adaptation scale to students of SMP YAPAN, Sawangan, Depok, West Java. The sampling technique used is purposive sampling, in which the criteria of research sampling has been done specifically with 154 respondents. To do the validity construct test on the performance anxiety instrument, the author uses CFA (confirmatory factor analysis) method. The logic of CFA according to Umar (2012) is:

- A. There is a concept or trait defined operationally so that questions or statements to measure it can be arranged. This trait is called factor while measurement on the factor is done through the analysis on the response of the items.
- B. It is understood that every item only measures one factor. It means, both item and subscale are unidimensional.
- C. The available data is used to estimate matrix, correlation between items that should be obtained, if the items truly are unidimensional. This correlation matrix is called sigma (Σ), then compared to the matrix from empirical data, called matrix S. If the theory is right, (unidimensional), there is no difference between matrix S – matrix Σ , or it can be stated with another statement such as $S - \Sigma = 0$.
- D. The above statement is used to make null hypothesis which is then tested using chi square. If the result is insignificant, $P\text{-value} > 0.05$, the null hypothesis is not rejected. In other words, the unidimensional theory, that one item only measures one factor, can be accepted.
- E. If the model fits, the next step is to test if the item is significant or insignificant to measure what should be measured using t-test. If the t-test result is insignificant ($\text{sig.} < 1,96$) the item, thus, is insignificant to measure what should be measured. The author then will have to drop the item.

- F. Next, if in the result of CFA, there is an item that is coefficient with the negative content, then the item should be dropped. It means the item measure things contradictory with what should be measured. However, the author will still need to recheck if the item is negative (unfavorable). Unfavorable item should be checked between CFA analysis is done.

III. RESULT

The following is the result of CFA (Confirmatory Factor Analysis) for every dimension:

A. Dimension of Cognitive

The author tests if the sixth item is unidimensional or measure cognitive aspect. The analysis result of CFA is done through one-factor model, which fits Chi-square = 22.83, df = 9, P-value = 0.00658, and RMSEA value = 0.101. Therefore, the author modified the model in which measurement errors on some items are allowed to correlate one another, resulting in the model that fits Chi-square= 9.26, df = 8, P-value = 0.32064, RMSEA = 0.032.

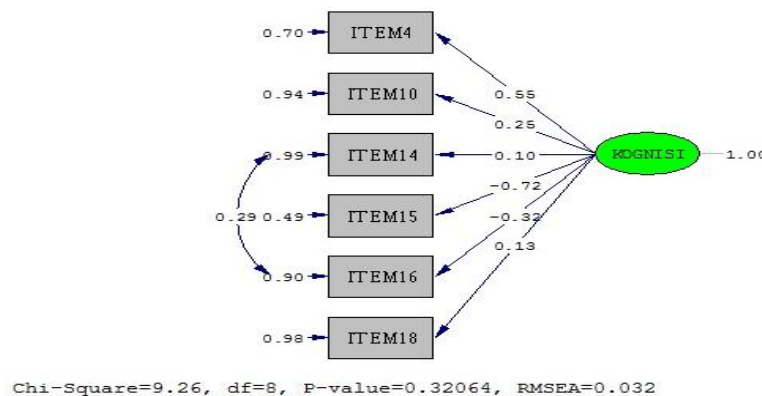


Figure 1. Cognitive Path Diagram

After the P-value > 0.05 is obtained, it can be concluded that the model with one factor can be accepted. It means all items only measure one factor: cognitive. The author investigated if the item measured the factor to be measured significantly and determined if the item should be dropped or not; the testing was done by looking at t value of every factor loading coefficient, such as in table 1 below.

Table 1. Item Factor Loading

Dimension	No. Item	Lambda	Std. error	T- Value	Information
Cognitive	4	0.55	0.12	4.73	Valid
	10	0.25	0.10	2.44	Valid
	14	0.10	0.10	0.97	Valid
	15	-0.72	0.13	-5.38	Invalid
	16	-0.32	0.10	-3.06	Invalid
	18	0.13	0.10	1.27	Valid
Physiology	1	0.91	0.07	12.30	Valid
	3	0.66	0.08	8.46	Valid
	6	0.48	0.08	5.71	Valid
	8	0.24	0.09	2.82	Valid
	9	0.70	0.08	8.95	Valid
	12	0.36	0.08	4.24	Valid
Behavior	2	0.26	0.10	2.65	Valid
	5	0.45	0.10	4.33	Valid
	7	0.47	0.10	4.55	Valid
	11	-0.36	0.11	-3.16	Invalid
	13	0.38	0.12	3.16	Valid
	17	0.29	0.12	2.43	Valid

Based on the table 1, the t-value for factor loading coefficient of all items is significant because $t > 1.96$ is $t < -1.96$. The author, then, investigated the factor loading of an item to see if it has negative loading or positive; the result was there was no negative factor loading in the item.

B. Dimension of Physiology

The author test if all the five items are unidimensional, which means it only measure physiological factor. Based on the analysis result of CFA conducted with one factor model, it was obtained that the items did not fit the Chi-square = 28.43, df = 9, P-value = 0.00081, and RMSEA value = 0.119. Therefore, the author modified the model, in which the measurement error on some items are allowed to correlate each other, which makes the model fit the Chi-square= 7.28, df = 7, P-value = 0.40031, RMSEA = 0.016.

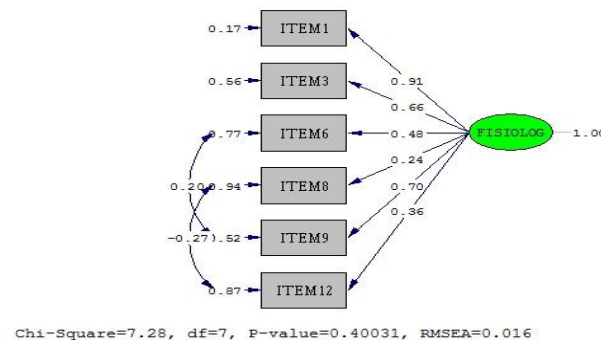


Figure 2. Physiological Path Diagram

Based on table 2, the t-value for factor loading coefficient of all items is significant because $t > 1.96$ or $t < -1.96$. Next, the author investigated the factor loading of the item to see if it has negative loading or not; it turns out to have no negative factor.

C. Dimension of Behavior

The author tested the five items to see if they are unidimensional or only measures behavior. From the CFA analysis result conducted on one factor model, the items did not fit the Chi-square = 31.24, df = 9, P-value = 0.00027, and RMSEA value = 0.127. Therefore, the author modified the model, in which measurement errors on some items are allowed to correlate each other resulting in models that fit the Chi-square= 10.21, df = 8, P-value = 0.25061, RMSEA = 0.043.

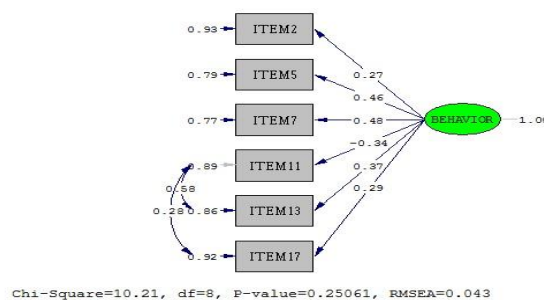


Figure 3. Behavior path diagram

Based on table 3, the t value for factor loading coefficient of all items is significant because $t > 1.96$ or $t < -1.96$. Next, the author investigated the factor loading of item to see if it has negative loading or not; it turns out that the item does not have negative loading.

IV. CONCLUSION

The construct validity test result on performance anxiety instrument using confirmatory factor analysis (CFA) approach revealed that all items are unidimensional, that is it only measures one factor. It can be concluded that one factor model theorized using performance anxiety instrument can be accepted excepts three items: item no 11, 15, and 16. Despite the three items, the overall

items can still be concluded as valid and unidimensional (only measures one factor). The criteria of good items are: (1) possessing positive factor, (2) valid (significant, $t > 1.96$ or $t < -1.96$); in other words, the item is unidimensional.

V. RECOMMENDATION

- A. It is recommended that validity test be conducted with more respondents at different school.
- B. It is recommended that a deeper research be conducted on the proper intervention method development on the students with anxiety symptoms identified by the measurement tool used in this article.

REFERENCES

- [1] Allen, G.J., Lerner, W.M., Hinrichsen, J.J. (1972). Study Behaviors and Their Relationships to Test Anxiety and Academic Performance. *Psychological Reports*, 30, 407-410.
- [2] Beaton, D.E., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). Guidelines for the process of cross-cultural adaptations of self-report measures. *Spine*, 25 (24), 3186-91.
- [3] Cassady, J.C., Johnson, R.E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, 27, 270-295.
- [4] Eysenck, M.W., Derakshan, N., Santos, R., Calvo, G.M. (2007). Anxiety and Cognitive Performance : Attentional Control Theory. *Journal of American Psychological Associations*, 2, 336-353.
- [5] Kenny, D.T., Davis, P., Oates, J. (2004). Music performance anxiety and occupational stress amongst opera chorus artists and their relationship with state and trait anxiety and perfectionism. *Anxiety Disorder*, 18, 757-777.
- [6] Leao, M. V. Coping with performance anxiety on teenage piano students. *Journal of Macao Polytechnic Institute*, 21, 35-47.
- [7] Umar, J. (2012). *Confirmatory Factor Analysis : Bahan Ajar Perkuliahan*. Jakarta: UIN Syarif Hidayatullah.
- [8] William, W.K., Zung. (1971). A rating instrument for anxiety disorders. *Psychosomatics*.

DEVELOPMENT HYPOTHETICAL MODEL RESOURCES MANAGEMENT STUDIES TEACHERS OF HINDU RELIGION

Aris Biantoro¹, I Made Sutharjana², WayanSukarlinawati³

¹ Lampung University, ² STAH Lampung, ³ STAH Lampung
biantoroaris@yahoo.co.id

Abstract- Learning at this time should use a variety of many resources, but in many cases there are many teachers do not use it. Especially teachers of Hinduism, for that we need the development of the management model of learning resources learning of Hinduism. Based on the characteristics of a Hindu religious education teachers of Hinduism should be able to manage learning and make learning more effective, efficient and has an appeal in order to achieve the learning objectives based on the characteristics of the field of study of the Hindu religion. Efforts are being made to acquire the learning outcomes and achievement of learning goals in accordance with the characteristics of Hindu religious education, through the development of learning resources management model.

Research methods This paper is produced from some theoretical studies, from some of these studies later produced the study variables. The study was conducted in three phases, namely: (1) a preliminary study, the assessment of the condition management of learning resources factual/existing and literature studies, (2) development of a prototype hypothetical model management learning resource teachers of Hinduism, and (3) the outcome of research, namely resulting final hypothetical model. Based on the results of the preliminary analysis it is found that there is no factual model of management learning resource teacher of religious studies Hindu primary school in Central Lampung. Based on the study of the learning technologies and learning management then the resulting hypothetical model management learning resource teachers of Hinduism.

Keywords: *hypothetical model, learning resource teachers of Hinduism*

I. PRELIMINARY

Learning as a human activity to construct or create knowledge by trying to give meaning to the knowledge in accordance with the experience of learners, Tasker (1992: 30), Wheatley (1991: 12), Tytler (1996: 20). According to Vygotsky (1990), and Hanbury (1996: 3), learning is a process that involves two important elements, namely the process of biological and psychosocial processes. By the time someone gets the stimulus from the environment, it will use the tools of his senses to capture or absorb the stimulus, then the brain nerve accepted information is processed. So that the learning process can be run with good results need to be developed based learning model various sources.

Learning the various sources is Encompassing a wide range of means by the which students are Able to learn in ways that are on a scale from Reviews those that are mediated by tutors to Reviews those where the students are learning independentl, (Brown & Smith, 1996). This will benefit the students to gain an understanding of the material in depth, dig up more information and produce learning outcomes. higher quality, improve the skills of divergent thinking, improve the skills of information processing, allowing the formation of knowledge in each phase, increasing positive attitude towards learning materials and academic achievement, making enthusiastic to learn and be inspired to actively participate and improve academic achievement, the planting of attitudes and critical thinking , (Sitepu, 2009).

According Seels and Richey (1994), learning resources are physical manifestations of technology hardware, software, and learning materials. Learning resources according Degeng (1990), are all sources that may be used by participants didiki learned behavior to occur, can be

used singly or in combination, either planned or learning resources learning resources utilized. According to the Association for Educational Communications and Technology learning resources is everything or power that can be utilized by teachers, either separately or in a combined form, for the sake of learning and teaching with the aim of increasing the effectiveness and efficiency of learning objectives. Learning resources can be grouped into two parts, namely: (a) a source of learning intentionally planned, (learning resources by design), ie all the sources that have been specifically developed as a component of instructional systems to provide learning facilities were targeted and formal; and (b) used to learning resources (learning resources by utilization), the learning resources that are not specifically designed for learning purposes yet to be found, applied and used for learning purposes, one of which is the mass media.

The use of a variety of learning resources in the learning by elementary school teachers intangible information / messages by design is learning materials, and by utilization is donggeng, poems, songs, and short stories. Conveys information design is dominated by teachers, whereas by utilization never used. Information storage media by design that is often used is the picture, and by utilization are newspapers, realia, and discarded packaging. Teachers do not use the equipment to move the message both by design and by utilization. How messaging by design very often used by teachers are lectures, discussion and debate, whereas by utilization is donggeng. Environment to deliver a message by design is very often used by teachers are classrooms, and libraries and laboratories that are not often used, as well as by utilization is very often used is the school and the surrounding neighborhood (Herpratiwi: 2015).

The importance of managing learning resources well in accordance with AECT (Association of Education and Communication Technology) (1977) defines the learning resources are various or all sources either in the form of data, people and a particular form that is used by students to learn, either separately or combined, making it easier students in achieving the learning objectives. Learning resources according to AECT divided into six types: Message (massage), person (person), material (material), Tool (Dvice), Technique (Technique), Environment (setting).

Based on the exposure shows that learning must now use a variety of learning resources, but on the field there are teachers who do not use it. Especially teachers of Hinduism, for that we need the development of the management model of learning resources in the learning of Hinduism.

II. RESEARCH METHODS

This paper is produced from some theoretical studies, from some of these studies later produced the study variables. The study was conducted in three phases, namely: (1) a preliminary study, the assessment of the condition management of learning resources factual / existing and literature studies, (2) development of a prototype hypothetical model management learning resource teachers of Hinduism, and (3) the outcome of research, namely produces a final hypothetical model (Herpratiwi: 2013).

III. RESULTS AND DISCUSSION

Management of Learning Resources (PSB) teachers of Hinduism is based on the theory of learning design, learning management theory. Instructional design theory is used to make decisions about the need for the management of learning resources so that learning can be an effective, efficient and attractive. The theory of learning management functions of planning, organizing, directing, controlling and evaluation of learning resources in accordance with the characteristics of Hindu Religious Education to be applied in the classroom. Based on this, the necessary learning resource management model, providing ease and assist teachers in performing their duties keprofesiannya and increased use of learning resources (Herpratiwi: 2013).

A. Model Development Learning Resource Management Master field of study Hinduism, the outline is done in three stages, namely:

1. Learning Resources Management Condition Assessment Factual / Existing and Literature (Pre Existing Assessment)

An assessment of the management of existing learning resources are focused on the aspects of learning resources used by teachers and management of learning resources. Assessment of the aspects of management (management) through interviews with the elements involved in the management of learning resources as respondents. Information obtained as follows:

- a. Aspects of learning resources used by teachers Types of learning resources that teachers use limited text books, textbooks, worksheets and myself as a model.
- b. Management of learning resources
Learning resources have not been managed, both in terms of planning, organizing, directing, controlling, and evaluation. Theoretical studies /literature study is a preliminary study of descriptive nature, which identifies how the hypothetical model management learning resource teachers of Hinduism developed, so it can be used as predictive, and consideration in the development of the model, which is ultimately to facilitate teachers in the use of learning resources and increase the student's character. Description of the results of theoretical studies were developed on the basis of learning theory behaviorism, cognitive social Vigotsky, information processing, learning theory Bruner, cybernetic theory, learning theory, Gagne, and Reigeluth, as well as management theory (Herpratiwi: 2013).

Besides the description of theoretical studies is also strengthened by a constructivist approach to learning in accordance with the purpose of teachers in the use of learning resources in the field of study of Hinduism. Behavioral learning theory emphasizes the conditioning factor in learning is the stimulus causing the response. In order for optimal learning outcomes, the stimulus should be designed in such a way and come from a variety of sources, so that the students responded easy. Social cognitive learning theory Vygotsky (1990) argues that learning occurs when students work or learn to handle the tasks that have not been studied yet those tasks are still within reach of his ability (zone of proximal development). Vygotsky also explains that learning occurs in two stages: The first stage occurs when collaborating with others / level social context, and the next phase is done individually, in which a process of internalization. During the process of interaction occurs, both between teachers / adults and students and among students, such as the ability to respect each other, to test the truth of statements of others, negotiate and adopt mutually opinions can flourish. The level of institutional and social context are interpersonal. It serves as scaffolding. Institutional interactions give children a behavioral norms and social comprehensive to guide his life. Interpersonal level has a more direct effect on children's mental functioning. According to Vygotsky (1990), skills in mental functioning develops through social interaction.

Bruner learning theory to explain how people learn or acquire knowledge and menstranformasi knowledge. Rationale The theory considers that human beings as processors, thinkers and creators of information. Bruner stated learning is an active process that allows people to discover new things beyond the information given to him. According to Bruner intellectual development of children follow a three-stage sequential representation, namely: a) enaktif, all the attention of the child depends on his response; b) the iconic, depending on the child's thinking pattern of sensory organization and c) the symbolic, the child has to have full understanding about something so that the child has been able to express his opinion with the language. Implications of the theory of Bruner in the learning process is exposes children to a confusing situation or a problem. With the experience of the child will try to adjust or reorganize the structures of ideas in order to achieve a balance in his mind. According to the cybernetic learning theory, learning is processing information, which explains the processing, storage, and recall of knowledge of the brain (Slavin, 2000). This theory explains how one can obtain a range of information and remember in a long time. Therefore it is necessary to apply a specific learning strategies that can facilitate all information is processed in the brain through multiple senses. Information will determine how the process or way of learning will take place. An information may be learned by a student with a kind of learning process, and the same information may be learned other students through different learning processes. According to Ausubel (1960) the acquisition of new knowledge is a function of cognitive structures that have been owned by individuals. Reigeluth and Stein (1983) says that knowledge arranged in hierarchical cognitive structure, knowledge of a more general and abstract obtained first by an individual to facilitate the acquisition of new knowledge that is more detailed. The implication in the learning, the better way of structuring knowledge as a knowledge base that came later, the easier the knowledge searchable and be recalled when needed. While the constructivist paradigm in learning means that students have the ability to explore and construct their own knowledge independently, Good & Brophy (in Kauchack&Eggen, 1998).

Piaget in Hans G. Furth (1970), asserts that knowledge is constructed in the minds of learners through assimilation and accommodation. Implications of the study are: (a) the language and

ways of thinking learners are different from adults, (b), the participants will think better if it can interact and deal with the environment well, (c) the materials studied learners should feel new but not foreign, (d) learners should be given the opportunity to learn what you want, talk and discuss with friends. One of the basic concepts of constructivism approach to learning is the social interaction of individuals with their environment. According to Vygotsky (1990), and Hanbury (1996), learning is a process that involves two important elements, namely the process of biological and psychosocial processes. By the time someone gets the stimulus from the environment, it will use the tools of his senses to capture or absorb the stimulus, then the brain nerve accepted information is processed. While constructivist paradigm according to Slavin (1994) is, (1) top-down processing, starting from complex problems to be solved and then generate or find the necessary skills, (2) cooperative learning, making the social environment and study groups as a place to gain knowledge, (3) generative learning, integrating the new knowledge schemata already owned. Based on the study above, Reigeluth (1983) developed a strategy of structuring the content or subject matter that relate to the four problem areas, namely; election (selection), the arrangement sequence (sequencing), summary (summary), and synthesis (synthesizing), namely: (1) if the subject matter content laid out by using a sequence from general to detail, the content or subject matter on a general level will be framework for linking the contents of other details. This is consistent with the structure of representation of information in the Long Term Memory (LTM), which will facilitate the search process back information. (2) If the summary is integrated into the strategy of structuring the content or subject matter, then it will function shows kapada students (the study) what information should be addressed in addition to saving capacity Working memory (WM). According to Gagne (1979), in the learning process occurs receipt of the information, to then be processed to produce output in the form of learning outcomes. Information processing occurs the interaction between the conditions of internal and external conditions of the individual. Internal conditions within the individual that is the condition necessary to achieve the learning outcomes and cognitive processes that occur in an individual. While external conditions is the stimulation of the environment that affect people in the learning process. According to Gagne stages of the learning process includes eight phases, namely, (1) motivation; (2) understanding; (3) acquisition; (4) storage; (5) recalling; (6) generalization; (7) treatment and (8) feedback.

In essence, learning is generally depicted Gagne (1979) in an effort whose goal is to help people learn. Events learning occurs when students actively interact with learning resources are regulated by the teacher. In the learning interaction, each student is treated as dignified humans, the interests and potential needs to be realized optimally. The quality of learning is strongly influenced by learning methods that do the strategy of macro and micro organizational learning, learning delivery strategy, and learning management strategies under the condition that there are characteristics of the destination, contents characteristics, constraints, and student characteristics. Learning outcomes can be classified into three, namely: (1) the effectiveness (effectiveness) (2) efficiency (efficiency), and the appeal (appeal). The effectiveness of learning, usually measured by the level of achievement si-learning, namely: (1) the accuracy mastery of learned behavior or often called the error rate (2) the speed of performance, (3) the transfer of learning, and (4) the level of retention of what studied. Learning efficiency, measured by the ratio between the effectiveness and the amount of time that students use and / or amount of the cost of learning used. The appeal of learning, measured by observing the tendency of students to keep / keep learning. The appeal of learning is closely related to the appeal of the field of study, where the quality of teaching will usually affect both.

1. Develop a prototype model of hypothetical Learning Resource Management (PSB) Master Field of Study Hinduism Based on the study of theory, then the resulting prototype hypothetical model PSB Master Hindu religious studies similar to the hypothetical model of character-based PSB according Herpratiwi (2013), namely:
 - a. Plan
PSB planning activities teachers of Hinduism include:
 - a) Determination of learning resources
Planning learning resources for learning formulated jointly between teachers and facilitators.
 - b) Management.
Management is done with the training of teachers in choosing the design and use of learning resources Hindu religious studies.
 - c) Management class

Classroom management is done by providing classrooms designed for group learning, and enable the mobility of students to work in groups. Each room is equipped class facilities and infrastructure include: the discussion table and chairs, LCD, Wifi internet, air conditioning, whiteboard, markers, and CCTV. With CCTV all activities can be monitored simultaneously.

- d) Management students
Students are given a briefing before the following study. Students are organized into groups consisting of 5 students per group.
- e) Organizing
Organizing by forming PSB unit manager with the following membership: the chairman / facilitator or coordinator in charge, observers and recorders.
- f) Directing class
Teachers practice activities PSB always got a briefing from the facilitator.
- g) Control
Control aims to harmonize between the facilitators and teachers, and post test of PSB Hindu religious studies.
- h) Evaluation
The evaluation was done both on students and teachers. The instrument used in the form of a questionnaire. Evaluation of the facilitator conducted through meeting forums. Aspects of the quality of learning resources (quality learning Recoursrch = QLR). Answer responen given in the form of a scale of 1 to 5 with the following conditions: STS = Strongly Always (score = 1), TS = Not always (score = 2), N = Neutral / Excellent (score = 3), S = Always (score = 4), SS = Very always (score = 5). To measure aspects of PSB Hindu religious studies, among others:

1. Aspects of quality of learning resources (Quality Learning Recoursrch = QLR)
Learning is a process that is fun students. With the management of learning resources students are given the opportunity to use a variety of learning resources to study according to the problem being studied. Thus, it will make students feel more relaxed and flexible. There are nine indicators to measure this aspect, as presented in Table 1.

Table 1. Aspects of quality of learning resources (Quality Learning Recoursrch = QLR)

No	Statement	Number
1.	Source learning by the teacher made me happy learning	1
2.	Sources of learning by the teacher allowed me to learn	1
3.	Many of the learning resources that must be sought to solve the problem, in addition to those already provided teachers	1
4.	Additional learning resources I was looking to solve the problem should be consulted prior to the teacher first	1
5.	We are given many choices of activities what should we do first	1
6.	We often given the opportunity to discuss with teachers with regard to learning resources that we use	1
7.	We are still struggling to understand the source of learning of the teacher	1
8.	Learning resources are provided teachers can develop interest in learning my	1
9.	In the study, there is the opportunity to choose the learning resources that will be studied	1
Total		9

3. The hypothetical model of PSB Master Hindu religious studies
Here is a model of a hypothetical PSB Master Hindu religious studies.

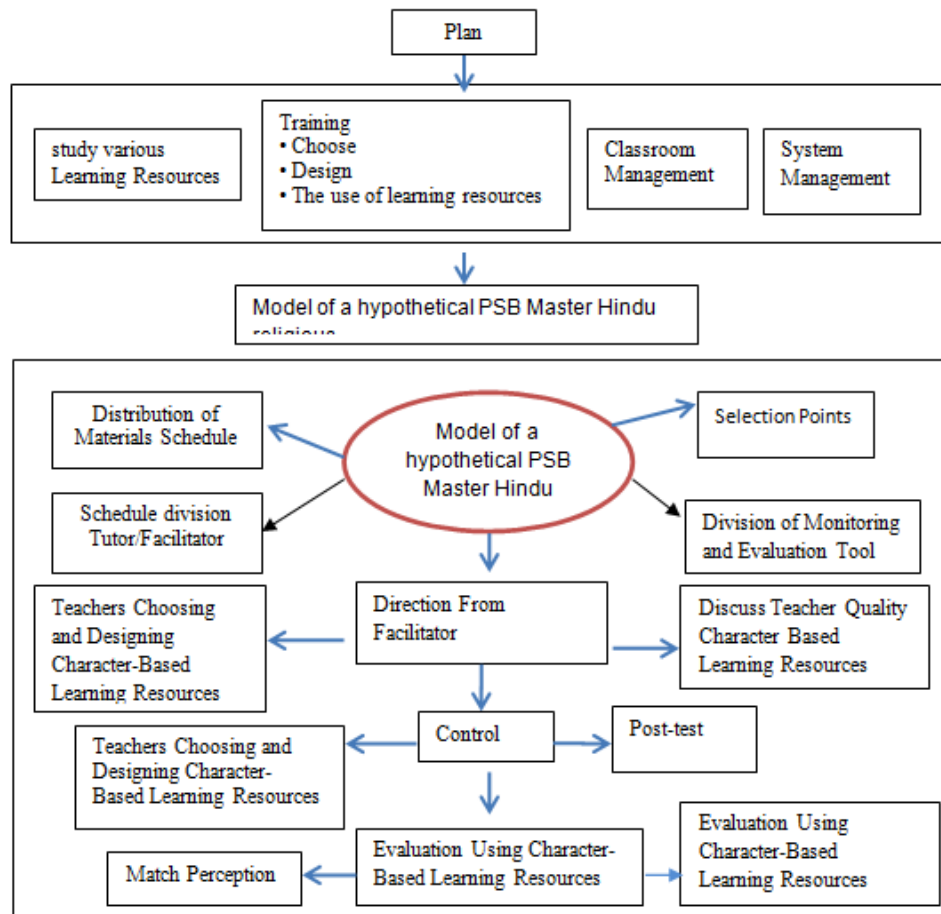


Figure 1. Model of a hypothetical PSB Master Hindu

IV. CONCLUSION

Based on the results of the preliminary analysis it is found that there is no factual model of management learning resource teacher of religious studies Hindu primary school in Central Lampung. Based on the study of the learning technologies and learning management then the resulting hypothetical model management learning resource teachers of Hinduism.

BLIOGRAPHY

- [1] AECT. 1977. Selecting Media for Learning. Washington DC: Association for Education Communication and Technology.
- [2] Ausubel, D.P1960. "The use of advanced organizersmin the learning and retention of verbal materialmeningful" Journal of educational psychology, 51. 267-272.
- [3] Brown, Sally & Brenda Smith. 1996. Resource Based Learning. London: Kogan Page Limited.
- [4] Degeng, I NyomanSudana. 1989. Teaching Science Taxonomy variable. Jakarta: Department of Education
- [5] Dick Walter and Lou Carey. 2005. The sistematic Design of Instruction. New York: Pearson.
- [6] Gagne, R.M., & Briggs, L.J., 1979. Principles of Instructional Design, New York: Holt, Rennerhart and Winston.
- [7] Hanbury. 1996. constructivism: So What? In J. Wakefield and Velardi L. (Eds.).Celeberating Mathematics Learning (pp.3-8). Melbourny: The Mathematical Association of Victoria.
- [8] Hans G. Furth. Piaget 1970. For Teacher.Englewood Cliffs N.J. Prentice Hall, Inc.
- [9] Herpratiwi. 2013
- [10] Herpratiwi. 2015. Penggunaan Sumber Belajar Dalam Pembelajaran Oleh Guru Sekolah Dasar Di Kota Bandar Lampung 2015. Prosiding Konferensi Ilmiah Tahunan. 319-325, ISBN: 978-602-71325-7-3. 2015
- [11] [Kauchack, D.P., &Eggen, P.D.1998. Learning and Teaching: Research-Based Methods. (3rdedition). Boston: Allyn and Bacon Office of the Basic Education Commission, 2004
- [12] [Reigeluth, C.M. 1983. Instructional Design Theories and Models: An Overview of Current Status Reviewstheir f. Hillsdale, New York: Lawrence Erlbaum Associates, 3-36.
- [13] Seels, Barbara B & Rita C.Richey.1994. Instructional Technology, Definitions and region, the translation. Jakarta: Printing Unit UNJ.

- [14] Sitepu, Star Peter. 2009. Decision-Making in the Development of Learning Resources. Jakarta: Jakarta State University.
- [15] Slavin, R.E. 1994. Educational Psychology: Theory and Practice (4th Edition). Boston: Allyn and Bacon.
- [16] Tasker, R. 1992. "Effective Teaching: What Can a Constructivist View of Learning Offer "In The Australian Science Teachers Journal 38 (1): 25-34.
- [17] Tytler, R. 1996. Constructivism and Conceptual Change Views of Learning in Science. In Khasanah use Of IPA. 1 (3): 4-20.
- [18] Wheatly, G.H. 1991. Constructivist Perspectives on Science and Mathematics Learning. Journal of Research in Science Teaching. New York: John Wiley & Sons, Inc. 35 (1). 9-21.
- [19] Vigotsky and Education. 1990. Instructional Implications and Applications of Sociocultural Psychology. Cambridge: New York.
- [20] Psychology. Cambridge: New York.

Indonesian Adaptation of Organizational Commitment Questionnaire from Meyer & Allen, 2004

Baqiyatul Auladiyah

Faculty of Psychology, State Islamic University Syarif Hidayatullah Jakarta
kiki_baqiyatul@yahoo.co.id

Abstract: The objective of this study was to examine the construct validity of Meyer and Allen (2004) organizational commitment questionnaire. The scale was distributed to 53 students in State Islamic University Syarif Hidayatullah Jakarta. The scale was categorized of three dimensions; affective commitment (AC), continuance commitment (CC) and normative commitment (NC). Each dimension consists of eight items, used an original version. The procedure of scale adaptation was conducted using the guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures (Beaton, 2000). We use a Confirmatory Analysis Factor (CFA) to analyze the construct validity. The result of this study was 22 from 24 items from this scale is valid ($t > 1.96$). It can be used for research. However, the two items of AC and CC must be dropped because of invalid.

keyword: *affective, continuance, normative commitment, construct validity.*

I. INTRODUCTION

Organizational commitment is often associated with feelings of individuals tied to the organization, so they are willing to take the time to think and act for the betterment of the organization though time consuming, mind and financially. Robbins (2001) explains that the organizational commitment is about employee side related to the organization in achieving organizational goals and always keeps its membership to the organization. Bateman and Strasser; 1984; Morris & Sherman; 1981; at Cohen (2003), explains that the level of performance of employees in an organization influenced by the employees organizational commitment. It will also affect employee productivity and low absenteeism and tardiness.

Mowday et al. (at Jex, 2002) explains that organizational commitment can be seen from the dedication of employees of the organization in accordance with his work, and work sincere with all heart and soul to achieve organizational goals. Meyer and Allen (2004) describe the organizational commitment as a sense of engagement so that they keep their membership in the organization. Therefore, organizational commitment can increase the stability of the organization as well as lower levels of turnover. More clearly, Meyer and Allen (2004) explained that employees who have a high level of commitment at the organization will work harder and more extras in achieving organizational goals.

The previous research explained that the commitment does not directly contribute to reducing turnover, but the organizational commitment give impact on employee performance, then influence the level of turn over itself (Tett & Meyer, 1993; Meyer, Stanley, Herscovitch, & Topolnytsky, 2002). Research on organizational commitment has been carried out by previous studies involving organizational culture variable, turn over, spirituality, work engagement, job satisfaction and many other variables.

The use of variable organizational commitment often associated with three dimensions of organizational commitment by Meyer and Allen (1990), namely affective commitment, continuance commitment, and normative commitment. The original version of each dimension consists of eight items so that the total amount to 24 items (Meyer & Allen, 1990). In 1993, Meyer, Allen and Smith (1993) have revised the items which originally totaled 24 items revised to 18 items with each dimension consists of six items. Meyer, Allen and Smith (1993) argued that there are some items that are multidimensional, not just measure one factor. According to Meyer and Allen (1990), affective commitment explained about an employee is emotionally attached, recognize and involved in the organization. Thus, employees who have a strong affective commitment will continue to work in the organization because they really *want to* do that. Meyer

and Allen (at Jex, 2002) explains that employees will tend to develop the type of affective commitment if their acceptance of the organization is fair in treating them.

Next, Becker (at Jex, 2002) explained continuance commitment with regard to the concept of side-bets orientation that emphasizes the contribution of a person who at any time can be lost if the person leaves the organization. This commitment is further explain about the employee perception of the loss he would face if he left the organization. Meyer and Allen (1990) suggested that employee with the high of continuance commitment, fuse the organization because they require the organization. This commitment is based on the rational needs. In other words, this commitment is formed on the basis of profit and loss, considered on what must be sacrificed when settling on the organization. The key to this commitment is a necessity for survival (need to).

Normative commitment can be explained with feeling such as loyalty, affection, warmth, ownership, pride, happiness, and others. Normative commitment is based on the obligation to work for the organization. Meyer and Allen (at Jex, 2002) also explains that employees who have a high normative commitment, remain members of the organization because they have to do. This commitment is based on the norms that exist within the employees, which contain an individual's belief will be the responsibility of the organization. They survive because of loyalty. The key to this commitment is the obligation to stay in an organization (ought to). This type of commitment is more due to the moral values that employee personally.

Research on organizational commitment validity and reliability of Meyer and Allen (1990) has been carried out by previous researchers. In this case, Meyer and Allen (2004) explain that the reliability of items based on three dimensions of organizational commitment can be relied upon. But Meyer and Allen stated that the validity of the test item organizational commitment remains to be done because not all items can be customized with a work culture that exist in each country. Testing the validity of measuring instruments adapted organizational commitment Meyer and Allen (2004) can be used to determine the level of organizational commitment of employees in an organization, whether in the field of education, social, industry and government, as an evaluation. Hopefully, the organizational commitment of employees in an organization becomes a major guideline in advancing an organization in achieving its goals.

II. METHOD

Data distribution in this study used questionnaires. The subject of this study are from the student of state Islamic university in Jakarta who are active in extra-campus organization, called X organization. The sampling technique which used in this study is convenience sampling, anyone who is willing to fill out a questionnaire, and suitable with the criteria of the population should to be respondent. In this study, authors distribute 100 questionnaires to members of the organization through its chairman, while questionnaires collected only 77. The 77 questionnaires that have collected, 24 questionnaires can't be processed in statistical calculations because it does not suitable with the criteria in the calculation. So that the subjects in this study was 53 activist students, that can be analyzed using CFA.

In this study, researchers obtained data by using data collection tools in the form of attitude scale, where the subjects are asked to respond the statements by choosing a specific choice by providing a checklist (√) the choice of answers. The scale that used in this study was taking Likert scale models. The statement of Likert scale models (1-5) is composed of a positive statement (favorable) and negative (unfavorable) that there are five alternative answers are provided; strongly agree, agree, neutral, disagree, and strongly disagree. The procedure of scale adaptation was conducted using The Guidelines for The Process of Cross-Cultural Adaptation of Self-Report Measures (Beaton, 2000).

Table 1. Organizational Commitment Scale Blueprint

No.	Dimension	Indicator	F	UF	Total
1	Affective commitment	Employees have emotional attach	1, 2	3, 6	4
		Employees feel bound by the organization	4, 5	7, 8	4
2	Continuance commitment	Employees feel a loss if they left from the organization	9, 11	10, 13	4
		Employees get benefit in the organization	12, 15, 16	14	4
3	Normative commitment	Employee awareness that the commitment to the organization is indeed supposed to do	17,19, 21	18	4
		Employees feel indebted to the organization	22, 24	20, 23	4
Total			12	12	24

III. CONSTRUCT VALIDITY

Testing construct validity here using logic Confirmatory Factor Analysis (CFA) by Umar (2010), namely: (a) There are concept or trait that defined operationally called factors, and measurement of these factors is done through an analysis of the items response. (b) Theoretically, each item only measure one factor alone, that means either items or subtests are unidimensional. (c) If the theory is correct (unidimensional) then of course there is no difference between the matrix Σ - S matrix or can be expressed by $\Sigma - S = 0$. (d) The statement made null hypothesis is then tested by chi square. If the results of chi-square is not significant ($P\text{-value} > 0.05$), the null hypothesis is accepted. This means that unidimensionalities of theory can be accepted that the item or sub-test instrument measures only one factor alone. (e) If the model fit, then the next step to test whether or not a significant item measure what is to be measured, using the t-value. If the t-value is not significant ($t > 1.96$), it is not a significant item in measuring what is to be measured, if necessary, such items in drop and vice versa. (f) Finally, if the results of the CFA are items load factor coefficient is negative, then the item should be dropped. The CFA analysis testing like this is done by using *LISREL software 8.70*.

A. Construct Validity Testing

1. Affective Commitment

Authors examine whether there are eight items that really are unidimensional, measuring affective commitment. From the analysis conducted by the CFA model of a factor obtained models that do not fit. Then, after the modification model fit with the values obtained Chi Square = 17:57; df= 18; P-value = 0.48453; RMSEA = 0.000. From the result of Chi-Square value known $P\text{-value} > 0.05$ (not significant), it's meaning that the items on the dimensions of affective commitment only measure one factor alone.

Validity of test results above, a model with only one factor can be accepted, which means that all the items shown to measure one aspect, namely variable affective commitment. However, in this measurement model, there are errors measurement on the items which are correlated, so that it can be concluded that some of the items on her actual multidimensional respectively.

From Table 2 it can be seen that all items have a coefficient of positively charged, but item 7 significant not because it has a value of $t = 1.17$ ($t < 1.96$), while the other items are significant ($t > 1.96$). This means that item number 7 does not measure what is to be measured, therefore, item number 7 have to be dropped and are not included in the calculation of the factor scores.

Table 2. AC Item Factor Loading

Item Number	Coefficient	Error Standard	T value	Significant
1	0.93	(0.20)	4.64	V
2	0.92	(0.20)	4.60	V
3	0.82	(0.20)	4.00	V
4	0.63	(0.21)	2.94	V
5	0.93	(0.20)	4.65	V
6	0.79	(0.21)	3.85	V
7	0.26	(0.22)	1.17	X
8	0.50	(0.22)	2.33	V

Explanation : sign V = significant ($t > 1.96$) ; X = not significant

2. Continuance Commitment

The analysis conducted by the CFA model of a factor obtained model does not fit. Then, after the modification model fit with the values obtained Chi Square = 13:40; df= 19; P-value = 0.81752; RMSEA = 0.000. From the result of Chi-Square value known $P\text{-value} > 0.05$ (not significant).

Then the results above, thus a model with only one factor can be accepted, which means that all the items shown to measure one thing only, namely continuance commitment. However, in the measurement model, there are error measurement on the items are correlated, so it can be concluded that the item is actually multidimensional on him respectively.

At table 3 can be seen that all significant items except item number 15 because it has a coefficient of negatively charged and has a value of $t = -0.18$ ($t < 1.96$), while the other items are

significant ($t > 1.96$). That is item number 15 does not measure what is to be measured, therefore, item number 15 must be dropped and not included in the calculation of the factor scores.

Table 3. CC Item Factor Loading

Item Number	Coefficient	Error Standard	T value	Significant
9	0.83	(0.19)	4.27	V
10	0.71	(0.20)	3.59	V
11	0.82	(0.19)	4.21	V
12	1.07	(0.18)	5.97	V
13	0.95	(0.19)	5.10	V
14	0.97	(0.19)	5.20	V
15	-0.04	(0.22)	-0.18	X
16	0.95	(0.19)	5.04	V

Explanation: sign V = significant ($t > 1.96$); X = not significant

3. c. Normative Commitment

The analysis conducted by the CFA model of a factor obtained models that do not fit. Then, after the modification model fit with the values obtained Chi Square = 13.08; df= 14; P-value = 0.52059; RMSEA = 0.000. Chi-Square value known P-value > 0.05 (not significant).

From the results of path diagram above, thus a model with only one factor can be accepted, which means that all the items shown to measure one thing only, namely the normative dimensions of commitment. However, in the measurement model, there are errors measurement on some of the items are correlated, so that it can be concluded that some of the items on her actual multidimensional respectively. From table 3.3, it can be seen that all items have a positive charge and significant coefficient ($t > 1.96$). This means that all items in this dimension have not to be dropped and should be included in the calculation of the factor scores.

Table 4. NC Item Factor Loading

Item Number	Coefficient	Error Standard	T value	Significant
17	0.41	(0.14)	2.92	V
18	0.54	(0.13)	4.02	V
19	0.93	(0.13)	7.37	V
20	0.39	(0.14)	2.72	V
21	0.72	(0.14)	5.11	V
22	0.57	(0.14)	4.00	V
23	0.51	(0.15)	3.46	V
24	0.78	(0.13)	6.00	V

Explanation: sign V = significant ($t > 1.96$); X = not significant

IV. CONCLUSION

In general, the results of this study indicate that the items on the scale of organizational commitment from Meyer and Allen (2004) actually measure what is to be measured, that only measure themselves each on the dimensions of organizational commitment, although some items were found still correlated between one item to another. Of the adaptation and translation back has also gone through the proper procedures, so that after the processing of the process is so long, the reader can easily understand the items that were distributed to the respondents as the subject of testing the validity of a scale of organizational commitment.

Furthermore, the results of testing the validity of this commitment measuring devices can be used as a reference for the reader to do some research if you want to use the same variables. In addition, this measure can also be used for companies or organizations to determine the level of organizational commitment of employees or members. From the above, thus the scale of organizational commitment can be used as guidelines in further research, both in academic and industry as well as the organization.

V. DISCUSSION

Validity of test result above, generally, a scale according to the organizational commitment from Meyer and Allen (2004) can be used for research. However, in testing the validity of the explanation above, item number 7 on the dimensions of affective commitment and item number 15 on the dimensions of continuance commitment declared invalid. It's actually been a concern of the authors alone since the beginning of the process of adaptation, item number 15 under discussion between the writer and lecturer and friends in the class in translational stages. So it is suggested for further research to not using item number 15 on the study or it could be reviewing the item number 15. In addition, since the number of respondents who are too few, expected in the research that will be done to increase the number of respondents so that the desired objectives can be achieved.

REFERENCE

- [1] Meyer and Allen. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology*. The British Psychological Society.
- [2] Meyer and Allen. (2004). *TCM employee commitment survey*. Academic Users Guide 2004. The University of Western Ontario.
- [3] Cohen, A. (2003). *Multiple commitments in the workplace*. New Jersey: Lawrence Erlbaum Associates Publisher.
- [4] Jex, Steve. (2002). *Organizational psychology; A scientist practitioner approach*. United States of America: John Wiley & Sons, Inc.
- [5] Luthans, Fred. (2008). *Organizational behavior*. United States: McGraw-Hill.
- [6] Robbins, Stephen. (2001). *Perilaku organisasi*. Gramedia: PT Indeks.
- [7] Umar, J. (2010). *Bahan penelitian analisis faktor*. Institute Assesmen Indonesia.

CREATIVITY PROBLEMS TEST FORM STUDENTS COMPLETE DESCRIPTION OF LEARNING CONNECTION WITH LEARNING OUTCOMES COUNTING MATHEMATICS IN PRIMARY

Darmiyati

FKIP Universitas Lambung Mangkurat
saniahareem@gmail.com

Abstract. This study aims to determine the creativity of students in completing the test questions in narrative form in relation to learning mathematics numeracy with learning outcomes in primary school. This research was conducted at the State Elementary School Cindai Alus 1 Banjar district academic year 2015-2016. The study lasted for three months, from January to March 2016. The method used in this study is a survey, with a correlation technique. The target population of this research is all students of State elementary school Cindai Alus 1 Banjar district. Population involved fourth grade students who are taking a math lesson. Total sample as a whole there are 60 students. Instruments to retrieve the data are in the form of tests and documentaries. Aspects of student creativity observed in this study include smoothness, fluidity, detail, and students authenticity in expressing ideas. The results showed that: (1) the significant value of creativity of the students completed the test questions in narrative form in counting is greater than α 0:05 namely 0176 and normally distributed; (2) the variance of the relationship between the creativity of students complete the test questions in narrative form in learn counting with mathematics learning outcomes expressed linearly with f arithmetic $< F$ table (1,432 $<$ 4:00); (3) there is a significant relationship between the creativity of the students completed the test questions in narrative form in learning to count with mathematics learning outcomes in primary school where r count $>$ r table (0.340 $>$ 0.254); and (4) the higher the level of creativity of students complete the test questions in narrative form in learning to count, the higher the result of learning math.

Keywords: Creativity, learning math, and learning outcomes.

I. PRELIMINARY

A. Background

Learning Mathematics is one of the basic skills students need to have an important role in everyday life in the community, not only in the field of exact sciences but also in another subjects mathematics is required. On the other hand students find it difficult to accept the math, they think math is very difficult when compared with another lessons. Lack of student mastering the math lesson, showed by the low student learning, the value of daily tests, the results of math test obtained by the students, as well as the complaints of parents who find it difficult to help their children learn math, as well as complaints from the public as well as the institution where the students math score still not optimal.

The description of the low quality of education, especially the results of the evaluation of a mathematic conducted by "The Third International Mathematics and Science Study-Report (TIMSS-R 2007), reported that the student's mastery in basic education in mathematics were surveyed internationally, Indonesia ranks 36th out of 49 countries, the value obtained an average of 397 from the average standard international value of 500 (R & D Ministry of Education and Culture, 2011: 1-4).

Based on these data we can say that the student's mastery of mathematical concepts in primary schools is weak, whereas students' mastery of math will be a tool that can help a variety of other subjects both from the same level of education or at a higher level. Since the 1975 curriculum until now, efforts to improve student learning outcomes in a variety of fields of study, especially mathematic developed continuously by the government or private sectors.

These efforts include the improvement of curriculum, provision of textbooks, work shop models and learning strategies for teachers of mathematics, completeness props, and the improvement of educational infrastructure. But the reality of the teaching of mathematics today still has weaknesses that led to the lack of student mastery of the subject matter provided. Efforts have been made to improve the quality of education has not been entirely successful as expected by the government, the community, parents and educators, even seen their weaknesses not only in the mastery of the subject matter but also there is a tendency of students who act negatively toward subjects math, even raised the term indicated on mathematical subjects namely "desperately".

Creativity or creative thinking as an indicator of improving the quality of Indonesian human resources, as stated in the vision of primary school to develop as people who are devoted to God Almighty, noble, healthy, faithful, capable, creative, independent, and become citizens democratic and accountable (Susanto, 2013: 70).

The ability to think creatively can encourage a person to be more productive and able to solve problems encountered in daily life so that he can live more independently in the community, the state, and nation. Creative thinking ability is very important, especially in the face of challenges and advances in science and technology very rapidly today. , Therefore education programs that are directed to develop creative thinking of students both within the cognitive, affective, and psychomotor. In fact not every school is able to create an environment that can develop students' creativity. So far, most schools are trained knowledge memory, logical thinking ability in the form of ability to find the most appropriate response to the problem given the available information.

Creativity of students in math still not been fully implemented by the teacher. Based on the observation of learning activities in primary schools in most schools had focused approach to learning in groups and is responsible for the tasks assigned by the teacher either LKS or EHS. Problem is given it is only on the material being taught and exemplified by the teacher is not in the form of problem-solving, not giving a chance to the students to be creative in solving problems or in answering the questions provided.

One early attempt to do to improve student success in math is to investigate the students' creativity to solve problems in the counting narrative form.

B. Problem Formulation

Whether there is a relationship between students' creativity to solve problems numeracy learning outcomes with math learning of fourth grader SDN Cindai Alus 1 Kab. Banjar?

C. Research purposes

This study aims to determine whether there is a relationship between students' creativity to solve problems in numeracy with their learning outcomes in math of fourth grader SDN Cindai Alus 1 Kab. Banjar.

II. LITERATURE

A. 2.1 Math Learning Outcomes

Higgins stated that mathematics is the study of the concept, as well as the symbol of the relationship between the two (Higgins, 1983: 4). More explicitly, Russel (Bell, 1978: 260) argues that "Mathematics is the queen of the sciences" mathematics is the queen and other sciences are her underling. There are four main objectives to be achieved in the subjects of mathematics, namely, facts, skills, concepts, and principles (Bell, 1981: 108-109).

Romiszowski give an opinion that learning outcomes is a behavior that is measured through tests of the subjects studied, including the knowledge and skills of program learning, knowledge demonstrated by the information stored in the mind, while the skill shown by action or reaction that a person to achieve the goal (Romiszowski, 1981: 241). To achieve the level of quality of learning outcomes for students who get satisfaction, then it must be a stable learning environment, working hard to learn every study the material being studied (Robert S. Feldman, 1989: 250).

The results of student learning in math is the student's mastery of the subject matter that has been given and measured by tests given by teachers covering material: addition, subtraction, multiplication, fractions and measurement.

Based on various opinions and thinking of experts that has been said above it can be concluded that, the result of learning mathematics is students' mastery of subject matter, previously disclosed gain learning experience of students through the teacher's test scores.

B. Nature of Creativity Students in Problem Solving Test Shape Description

Creativity is a multi-dimensional concept, many definitions of creativity expressed by the experts. They review different aspects from the viewpoint each of the experts. (Santrock, 1997: 309) stated "Creativity is the ability to think about something in novel and unusual ways and to come up with unique solutions to problems".

Rhodes (Ali and Asrori, 2011: 44) suggests four dimensions in the definition of creativity is called the "Four P's of Creativity: person, process, product and press. In terms of personal creativity. If we review the concept of 4-P of personal creativity, the driving, the process and the product, it can be seen that the four concepts of creativity are interrelated and mutually.

Supriadi (Riyanto, 2012: 229) says that "The characteristics of creativity can be divided into cognitive and non-cognitive traits. Into the cognitive traits including four characteristics of creative thinking, namely originality, flexibility, smoothness, and elaboration. Into the non-cognitive traits including further motivation to say that non-cognitive traits, as important as cognitive traits because without support by the appropriate personality, one's creativity can not develop naturally. Furthermore, each of these aspects are outlined below.

Smoothness (Fluency) is the creativity aspect relating to the ability to produce or express ideas quickly to solve a problem, which is the main concern in this case is the number of ideas as a reaction of the problems encountered. The emphasis in terms of quantity rather than quality of ideas generated. However the relevance of these ideas with issues remain to be addressed. The more ideas generated individually, the person is more creative.

Dexterity (Flexibility) is an aspect of creativity that relate to an individual's ability to produce a variety of ideas in response to a specific problem or situation. In this case the main priority is the many different opinions that generated someone in response to an issue. Flexibility not only on different ideas, but also pointed to the numerous approaches in solving problems that can be chosen.

Detail (elaboration) is an aspect of creativity associated with the process of detailing the contents and ideas. In granting these contents, a person is required to give a description or explanation as detailed as possible to the ideas put forward, so it is easy to understand other people.

Authenticity (originality) is creativity aspects related to the ability of a person expressed his opinion that is unique and is statistically rare, meant here is the infrequency turnout compared with other ideas from friends in response to the same problem.

Creativity of students to solve numeracy problems is the ability of students to think of something new and unique to the unusual way he did to solve math problems that have been granted covering material as addition, subtraction, multiplication, and division. This ability also reflects the characteristics of creativity that includes fluency, flexibility, originality and detail.

C. Formulation Research Hypothesis

Based on the theoretical description that has been stated previously, then formulated the hypothesis of the study was a positive relationship between students' creativity to solve problems numeracy learning outcomes math fourth grader SDN Cindai Alus 1 Kab. Banjar.

III. RESEARCH METHODOLOGY

A. Population and Sample Research

This study was conducted at SDN Cindai Alus I class IV Banjar district from January to March 2016. The target population were performed in this study were all students of SD Negeri Regional District Banjar 2015/2016 academic year, while the population is the whole fourth grader students of Cindai Alus I, totaling 60 people. The sampling technique in this research is purposive random sampling.

The method used in this study is a survey, with a correlation technique. It is designed to obtain information about the symptoms at the time of the study. Survey methods may be collected information about the creativity of students with mathematics learning outcomes.

Data collection techniques used were test engineering research, and documentaries. The test is used to collect data about the creativity of students completing the counting (X), while data on student learning outcomes in math (Y) used documentary techniques namely by taking the value of tests that have been given a mathematics teacher in the first half.

B. Data Collection Instrument

Instruments creativity of students to arithmetic (X) is collected through creativity tests, which measured aspects include fluency, flexibility, detail and originality, with a sub-test as addition, subtraction, multiplication, and division. These tests form an essay test (description), and the number was around 14 questions. Every item students answered was appreciated and rated according to difficulty level items.

C. Data Analysis Techniques

Data obtained from this study were analyzed using descriptive statistics of frequency tables, the average percentage, and standard deviation. While inferensial statistics are used according to the required test, in this case the test for normality, linearity, and hypothesis testing research on each variable.

IV. RESULTS AND DISCUSSION

A. Results from Pilot Test Student Creativity Problem Solving Counting

Instrument before being given to the respondent, it must first be validated by three assessors refer to the grille that has been made in accordance with the fourth grader elementary school curriculum. The instrument contains four criteria used as guidelines to assess the validity of the creativity concept test students in solving arithmetic. Four criteria are (1) the suitability of the test items with the content or subject matter, (2) the suitability of number of questions with instructional objectives, (3) the suitability of language used, and (4) the construction of the tests used. If an item meets the four criteria, then given a score of 4, if three criteria were met by a score of 3, if two criteria were met by a score of 2 and if the criteria were met by a score of 1. Scores used is the average ratings of the three ratings.

The average yield instrument ratings turned out to test the creativity is eligible for tested in the study, although there are some things that need to be corrected, such as language. The results of the test instrument tests the creativity of students solving math has high reliability. This is shown also by the score of 0.88 and r of 14 test items are all qualified to serve as an instrument of research. This is demonstrated by the significant r each item.

B. Calculation Results Descriptive Analysis Creativity Students with Mathematics Learning Outcomes

Based on calculations show that for the creativity of the students obtained the maximum score of 28 and a minimum score of 12, an average of 19.53, and a standard deviation of 3.337.

C. Testing requirements analysis

1. Normality Test

Normality test data using test statistic K-S or Kolmogorov-Smirnov test. To determine the normality of the data using the comparison value Asymp. Sig.2-tailed at the alpha level of 0.05 (Syarifudin, 2010: 143).

The test results can be presented as follows:

Table 1. Results of Normality Test Data through Test One-Sample Kolmogorov-Smirnov Test

		Kreativitas Siswa	Hasil Belajar Matematika
N		60	60
Normal Parameters ^{a,b}	Mean	19.53	63.50
	Std. Deviation	3.337	8.991
Most Extreme Differences	Absolute	0.103	0.068
	Positive	0.089	0.068
	Negative	-0.103	-0.056
Test Statistic		0.103	0.068
Asymp. Sig. (2-tailed)		0.176c	0.200c,d

From Table 1 above it can be seen Asymp value. Variable-tailed Sig.2 Creativity, namely 0176 and 0200 Variable learning outcomes. Both values Asymp. Sig.2-tailed for each of the variable value is greater than the alpha level of 0.05, it can be concluded that the data come from populations with normal distribution.

2. Test Linearity

Linearity test according by Nisfiannor (2009: 92) is performed to determine whether the relationship between the dependent and independent variables is linear (straight line). Linearity test performed by test F. If the value of F is smaller than the F table at significance level of 5% as well as the significance value greater than 5%, then the data will have a linear relationship (Syarifudin, 2010: 150).

Table 2. Data Linearity Test Results

		Kreativitas Siswa	Hasil Belajar Matematika
Kreativitas	Pearson Correlation	1	0.340**
	Sig. (2-tailed)		0.008
	N	60	60
Hasil	Pearson Correlation	0.340**	1
	Sig. (2-tailed)	0.008	
	N	60	60

Based on Table 2 above it can be concluded the relationship between creativity and the students' mathematics learning outcomes which $F_{hitung} < F_{table}$ ($1.432 < 4.00$), the variance of the relationship between the creativity of students complete the test questions in narrative form and the learning outcomes stated linear mathematics.

3. 4.3.3 Hypothesis Testing

Hypothesis testing using Pearson Product Moment correlation variable counting the creativity of students to solve problems with mathematics learning outcomes. The hypothesis put forward are as follows:

Ho: There is no relationship between the creativity of the students completed the test questions in narrative form with the results of learning mathematics in SDN Cindai Alus 1 Kab. Banjar.

Ha: There is a relationship between the creativity of the students completed the test questions in narrative form with the results of learning mathematics in SDN Cindai Alus 1 Kab. Banjar.

Criteria: Reject H0 if Sig. (P) < 0.05

Significance: $r_{hitung} > r_{tabel}$

Results of correlation analysis using SPSS 23.0 is as follows:

Table 3. Results of correlation test between the creativity of the students completed the test questions in narrative form with the results of learning mathematics in SDN Cindai Alus 1 Kab. Banjar

		Sum of Squares	df	Mean Square	F	Sig.
Hasil Belajar * Kreativitas Siswa	(Combined)	1785.958	14	127.568	1.924	0.049
	Between Groups	551.660	1	551.660	8.322	0.006
	Linearity	1234.297	13	94.946	1.432	0.182
	Deviation from Linearity					
	Within Groups	2983.042	45	66.290		
Total		4769.000	59			

Based on Table 3 obtained by the Sig. 0.008 < 0.05, so that when incorporated into the criteria for testing the hypothesis, it can be concluded that the variable creativity of students to the learning outcomes of mathematical relationships or in other words H0 rejected and Ha accepted.

Meanwhile, to see the level of relations between the two variables can be obtained by comparing the value of r tables that can be obtained through the table with the number r Product Moment $df = 60$ ($df = n-1$) at the 5% significance level that is equal to 0.254. Therefore the value of r count $> r$ table ($0.340 > 0.254$), then the degree of correlation of the variable creativity to learning outcomes is significant.

Based on the calculation of descriptive statistical analysis that the level of students' creativity in solving arithmetic in SDN Cindai Alus 1 Kab. Banjar medium category with a score of 16.19 to 22.33 range. At this level caused by the writer's observation of the school in general are less trained in creating and developing the creativity of the students especially in terms of smoothness and detail put forward ideas.

The calculation result of research shows that learning outcomes math scores of fourth grader students of SDN Cindai Alus 1 is quite satisfactory with a percentage of 70% and a score of 54.509 -71.991 range. This is partly because different students' ability to understand the material given, learning strategies adopted by teachers in the classroom, and a student's readiness in accepting the given subject matter. Creativity is closely associated with students' attitudes and responsibilities in implementing and completing tasks, because without it the student can not achieve a high learning though potentially he could. The emergence of creative students preceded by a creative teacher anyway.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

Based on the research that has been done, it can be concluded that there is a relationship between students' creativity to solve problems counting the results of students' mathematics learning Elementary School Cindai Alus 1 in Banjar, these findings suggest that the higher the students' creativity to solve problems counting the results of student learning in math tends to be higher as well. Conversely, the lower the students' creativity to solve problems counting the results of student learning in math tends to decline.

B. Suggestions

First, students need to be aware of their roles and responsibilities as students who want to have the knowledge in mathematics, where mathematics covers a lot of material that requires a prerequisite knowledge and demands a different way of learning with other subjects. Students do not feel satisfied if you get a score of 70, students should learn from the mistakes committed and immediately addressed so it does not have problems in the future. Daily tests and test each lesson topic, which is often the teacher, whatever the result should not be interpreted simply to measure the level of mastery of the material being studied, but also look for the weaknesses and shortcomings that owned to be addressed immediately, to get optimal learning results.

Second, teachers should regularly give a test, not only for measuring the success of students but to identify the weaknesses and constraints experienced by students in the learning process. It is expected the students' creativity can thrive on the student's answers and can estimate the features and characteristics of the students and should be established good cooperation between parents and teachers, in particular in order to improve creativity and math student learning outcomes.

Third, the Institute of Education Workforce Education as PGSD as an education institution produce prospective elementary school teachers, the need for the establishment of communication skills early and integrated with program activities field experience and simulations or exercises taught with peers, in the implementation of subjects concerning the learning process, to evaluation system.

REFERENCE LIBRARY

- [1] Anwar Syarifudin, MA. 2010. *Metode Penelitian*. Yogyakarta: Pustaka Pelajar.
- [2] Balitbang Kemdikbud. 2011. *TIMSS (Trends In International Mathematics and Science Study)*. <http://Litbang.komdikbud.go.id/index.php/timss>. Tanggal 6 Maret 2016
- [3] Bell, Frederick H. 1981. *Teaching and Learning Mathematics in Scondery School*. New York: Printed in the United States of Amerika.
- [4] Feldman, Robbert S. 1989. *Essentials of Understanding Psychology*. New York: McCraw-Hill Book Company.
- [5] Nisfiannor, Muhammad. 2009. *Pendekatan Statistika Modern untuk Ilmu Sosial*. Jakarta: Selemba Humanika.
- [6] Riyanto, Yatim. 2012. *Paradigma Baru Pembelajaran*. Jakarta: Kencana.

- [7] Romiszowski, A. J. 1981. *Designing Intructional System*. New York: Kogan Page.
- [8] Santrock, John W. 1997. *Life Span Development*. Texas: Brown & Benchmark.
- [9] Susanto, Ahmad. 2013. *Teori Belajar dan Pembelajaran di Sekolah Dasar*. Jakarta: Kencana.

Effectiveness Guided Discovery Approachment Through Cooperative Learning Think Pair Share (TPS) Type in Terms of Students' High Order Thinking Skill (HOTS)

Deny Sutrisno¹.

¹(Farmasi, STIKES Harapan Ibu Jambi)
denysutrisno@gmail.com

Abstract—This study aims to describe the effectiveness of guided discovery approachment through cooperative learning Think Pair Share (TPS) type in terms of Students' High Order Thinking Skill (HOTS). This study was a quasi-experimental study using the pretest-post-test design. The research population comprised all Year X students of MAN Yogyakarta 3 and the sample was 1 out of 5 classes. The data collecting instruments consisted of a HOTS test (reliability 0,809). To describe the effectiveness of guided discovery approachment through cooperative learning TPS in terms of HOTS, the data were analyzed using the paired sample t-test at the significance level of 5%. The results show guided discovery approachment through cooperative setting TPS type are effective in terms of HOTS.

Keywords: *guided discovery, high order thinking skill, think pair share.*

I. INTRODUCTION

In order to confront AEC (ASEAN Economic Community), people demand on the quality and relevance of higher education. The quality of education was determined by several things, one of which was students' achievement. According to Brookhart [1] HOTS is associated with increased student achievement.

The guided discovery approachment through cooperative learning can be the solution to increase HOTS. Advantage of guided discovery by Westwood [2], the given topics are usually intrinsically motivating. Also according to Jolliffe [3] cooperative learning could be a useful strategy to support HOTS.

Guided discovery approach will be effective if organized into cooperative learning suitable for this approach. Because not of all cooperative learning model will be suitable to be combined with guided discovery. Guided discovery approach is an approach that has the basic activities in the form of discovery so it needed a model that supports these activities. Among the types of cooperative learning model that is suitable for guided discovery approach is Think Pair Share (TPS) because this type has activities that provide opportunities for students to discover something.

The result of discussion with Mr Suwandi, M.Pd, teacher of mathematic at MAN Yogyakarta 3, obtained information that students have difficulty in understanding the lesson. Students also having difficulty in analyzing the problem. Though the analysis is one of the levels in HOTS.

According of that, researcher wanted to know the answer from question how the effectiveness of guided discovery approachment through cooperative learning TPS type in terms of Students' HOTS. This study aims to describe the effectiveness of guided discovery approachment through cooperative learning TPS type in terms of Students' HOTS.

The significance of this study is expected to contribute positively to teachers, especially related to their capacity as educators, so that students will be have development in learning. In addition, this study is expected to be further study materials for researchers in similar research.

II. RESEARCH METHOD

This study was a quasi-experimental study using the pretest-post-test design. The research population comprised all Year X students of MAN Yogyakarta 3 in the academic year of

2014/2015 and the sample was 1 out of 5 classes. The respondents were students enrolled in X MIA 3. A total of 31 students took part in the study.

The data collecting instruments consisted of a HOTS test. Table 1 presents basic competencies of learning and indicator HOTS pretest-post-test item. This test include geometry subject consisting of five questions.

To proof validity of instruments in this study, content validity was used. In validating the test questions, the researchers asked the opinion of two lecturers of mathematics education Yogyakarta State University who are competent. Once the instrument is corrected by a validator, the instrument is then revised based on feedback given validator.

To estimate reliability of HOTS test instruments, Cronbach Alpha formula was used. The reliability coefficient calculation results of HOTS test instruments using Cronbach Alpha formula was 0.809.

Table 1. Basic Competencies And Indicator Hots Pre-Test And Post-Test Item

Basic competencies	Indicator
3.13 Describing the concept of distances and angles between points, lines and planes through the demonstration using model or other learning media.	1. Analyze, able to differentiate, organize and attribute. 2. Evaluate, able to check and critique.
4.13 Using the geometry principles in plane and space in solving real problems that relating to the distances and angles between points, lines and planes.	3. Create, able to generate, plan and produce.

The HOTS pre-test were administered in group testing sessions to students before they were given the guided discovery approachment through cooperative learning TPS type. Then after they were given the guided discovery approachment through cooperative learning TPS type, the HOTS post-test were administered. The results obtained converted into a value ranging between 0 to 100.

To describe the effectiveness of guided discovery approachment through cooperative learning TPS type in terms of Students' HOTS, the data were analyzed using the paired t-test at the significance level of 5%. The paired t-test used with assumption the data distribution was normal and homogen.

III. RESEARCH FINDING AND DISCUSSION

The results of HOTS pre-test and post-test are reported in Table 2. The data distribution was normal and homogen. To describe normality data, Kolmogorov-Smirnov test was used. To describe homogeneity data, Box's M test was used.

Paired-t test was conducted to examine effectiveness of guided discovery approachment through cooperative learning TPS type in terms of Students' HOTS. The test shows that guided discovery approachment through cooperative learning TPS type have significant influence to students' HOTS as indicated by < 0.05 level of significance.

Table 2. Result Of Hots Pre-Test And Post-Test

	TPS	
	Pre-test	Post-test
Means	4,63	35,87
Standard Deviations	4,60	19,93
Highest Score	13,00	67,00
Lowest Score	0,00	0,00
Total Students	31	31

The study shows that guided discovery approachment through cooperative setting TPS type are effective in terms of HOTS. This is consistent with the study of Asrul Karim [4] also identified guided discovery approachment can increase students' HOTS.

Based on cognitive levels of Bloom's Taxonomy revised Anderson & Krathwohl [5], Levels of cognitive taxonomy has six levels of cognitive processes with different levels of complexity. Six levels of cognitive taxonomy are remembering, understanding, applying, analyzing, evaluating and creating. The Higher level of thinking skill are analyzing, evaluating and creating.

Working cooperatively can help the development of high-level thinking skills such as analysis, synthesis and evaluation. Kagan [6] said TPS functions in academic and social are to improve ability of generating and revising hypotheses, inductive reasoning, deductive reasoning, application and also students' participation and involvement in learning. This was proofed, where students could complete the given problems on the worksheet after discussing with their pairing friend and also with the guidance of teachers.

TPS learning procedure starts with the teacher asking a problem. Then the students to think about the problems from the teacher (think). After that the students discuss in pairs (pair). Then the students share the results of their discussions with classmates (share). In the process of TPS, teachers can provide the necessary guidance so that students can find the right concept or principle. To help guide students in discovery process, teacher could use the worksheets.

This study show that TPS is matching with guided discovery approachment. TPS have discovery proses in it proses so we can get better benefit from combination of them. Westwood [2] said the benefits of guided discovery of which students are actively involved in the learning process and the topics are usually intrinsically motivated and also used in the context of the present invention gives the impression is often more than studying textbooks and exercise as usual.

In previous research, Marlina, et al [7] also identified learning model TPS type can improve students' mathematical disposition because students are actively involved in learning. It is reasonable because through TPS learning, student should listen to question, have time to think individually, talk with each other in pairs and finally share respons [8].

This study has limitations, which is expected to open up opportunities for other researchers to conduct similar research. The limitations this study were items instrument used in this study was less representative and less proportionately on all the indicators. Also, the questions used in the pretest and posttest considered less realistic.

IV. CONCLUTIONS AND SUGGESTIONS

In conclusion, the findings of this study show guided discovery approachment through cooperative setting TPS type are effective in terms of HOTS. The implication of this research, the researcher recommended to applied guided discovery approachment through cooperative setting TPS type to increase HOTS.

In future research, it will be important to determine whether guided discovery approachment through cooperative setting TPS type are effective in terms of other affective domain beside HOTS, which are competence that have consequences for intellectual enrichment and growth.

REFERENCES

- [1] Brookhart, S. M. How to asses higher-order thinking skills in your classroom, Alexandria: ASCD, 2010.
- [2] Westwood, P. What teacher need to know about teaching methods, Camberwell: ACER Press, 2008.
- [3] Jolliffe, W. Cooperative learning in the classroom: putting it into practice, London: Paul Chapman Pulishing, 2007.
- [4] Karim, A. Penerapan metode penemuan terbimbing dalam pembelajaran matematika untuk meningkatkan pemahaman konsep dan kemampuan berpikir kritis siswa sekolah dasar. Jurnal Penelitian Pendidikan SPS UPI, vol. 1. Pp. 21-32. 2011.
- [5] Anderson, L. W. & Krathwohl, D. R. A taxonomy for learning, teaching and assessing. New York: Longman, 2001.
- [6] Kagan, S. The structural approach to cooperative learning. Educational Leadership, vol. 47, pp. 12-15. 1989.
- [7] Marlina, Hajidin & Ikhsan. Penggunaan Model pembelajaran Kooperatif Tipe Think-Pair-Share (TPS) untuk Meningkatkan Kemampuan Komunikasi dan Disposisi Matematis Siswa di SMA Negeri 1 Bireuen. Jurnal Didaktik Matematika, vol 1, pp. 83-95. 2014
- [8] McTighe, J. & Lyman, F. T., Jr. Cueing thinking in the classroom: the promise of theory embedded tools. Educational Leadership, vol. 45, pp. 18-24. April 1988.

Indonesian Adaptation on Scale of Readiness for Organizational Change

Dharan Atasya Rakhmat

Faculty of Psychology, State Islamic University Syarif Hidayatullah Jakarta
dharanatasya @gmail.com

Abstract- The research was to examine the construct validity of scale adaptation on Holt et al. (2007) Readiness for Organizational Change. The scale consist of four dimensions; appropriateness, management support, change efficacy, and personally beneficial (personal valence). Process adaptation was conducted using the Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures (Beaton et al., 2000). Used a construct validity using Confirmatory Analysis Factor (CFA) on 400 employees of Pengelola Operasi & Pemeliharaan (OP) in Jatiluhur using convenience sampling technique. The result of this research was 18 of 41 items was adapted is valid ($t > 1.96$), which are 7 items of Appropriateness, 4 items of Management support, 3 items of Change efficacy, and 4 items of Personally Beneficial (personal valence). However, there is several items that should be dropped because of invalid; The conclusion of this research that this scale could be used and as an instrument for other researchers if they will measure a Readiness Change.

Keyword: *Readiness for Change, Adaption Scale, Confirmatory Analysis Factor .*

I. INTRODUCTION

In practice, the issue of changes in the organization have a negative impact, and certain anxiety because it affects the comfort zone of the employees. The negative impact caused a response in the form of resistance to change (Saragih et al., 2013), while the employee is the key to the successful implementation of the changes that need to be handled properly so that employees feel ready, and respond to changes as a challenge, so that changes can be successfully implemented (Zulkarnain & Hadiyan, 2014).

Resistance to change also occurred because employees perceive the change as a condition that threatens his position (Proctor & Doukakis, 2003), and created uncertainty about the future (Elving, 2005). Uncertainty is also information obtained from the obscurity of the organization (Barrett, 2002). Ideally leadership as an arm of the organization is able to communicate, and educate his subordinates to support the organization's strategy, so that the resistance of employees on the changes can be handled or minimized, and make the changes effective implementation (Mangundjaya, 2012; Holt et al., 2007).

A. Readiness for Change Measurement

To measure the readiness of the organization in the face of change, can use several approaches. Below will be described the readiness measurement changes that popular used in various studies, namely:

1. Readiness for Organizational Change Scale (Hanpachern et al., 1998)

Hanpachern et al. (1998) evaluated the process undertaken by the leadership to measure perceptions and opinions of employees in the proposed changes. Employees believed to have a willingness to support and participate in the change to evaluate the leadership in communicating changes to their subordinates. This instrument consists of 14 questions divided into three domains, namely:

- a. Participating (participating in the process of change)
- b. Promoting (supports the changes)
- c. Resisting (resistant to change)

2. Readiness for Organizational Change Scale (Holt et al., 2007)

This approach is a measurement that was developed and evaluated, so that it can be used to measure Readiness for organizational change on an individual level as activity

changes initiated and carried out individually in the organization. The main objective is to propose a quantitative measurement at the individual level that meets the requirements of psychometric (American Psychological Association [APA], 1995), and can be used in the system changes that can touch many aspects of the organization. This questionnaire consists of 44 questions with four dimensions measuring as follows:

a. appropriateness

Dimensions appropriateness explain the individual's belief, that a proposed change is right for the organization, and the organization will benefit from these changes.

b. Management Support

Dimension Management Support explained the beliefs or perceptions of individuals that leadership and management will support and commit to the planned changes.

c. Change Efficacy

Dimensional Change Efficacy explain the individual's beliefs about the ability to implement the desired changes (skills, and able to carry out tasks related to organizational change).

d. Beneficially Personal (Personal Valence)

Personally Beneficial Dimensions (Personal Valence) explains the benefits that can be felt personally if the change is implemented.

3. Readiness for Organizational Change Scale (Bouckennooghe et al., 2009)

This approach is a measurement that was developed by considering several factors that hinder the readiness for change, that climate of change, and how organizations cope with his change process (process of change). Here are the dimensions of the measurement:

a. Internal concept: (1) Trust in Leadership is how subordinate supervisors and top-management perceives as a party that can be trusted (practice what they are saying, promise, honest and fair, giving the feeling of safety in presenting ideas), (2) Politicking explain the political game level within the organization. High political intensity leads to unnecessary costs, cause delays, and the reluctance of sharing knowledge, (3) Cohesion refers to the role of cooperation and trust of employees of competence (unity in the organization, willingness to support colleagues).

b. Process: (1) Participation is the extent to which staff are involved and convey decisions that directly impact on employees (arrangement of bottom up procedure, etc.), (2) Support by supervisors understood as the extent to which employees have the support and understanding of the supervisor on it (measure of openness of employees, and the ability to direct subordinates to pass through the process of change), (3) Quality of change communication refers to how changes delivered (effective communication: clarity, frequency, and openness); Attitude of top management toward change involves mental taken from top management-related changes (if capable of supporting change initiatives or whether actively involved in the change?) .

c. Readiness for Change: (1) Emotional readiness for change is affective reactions to changes, (2) Cognitive readiness for change are the beliefs and thoughts held beliefs about the change (change benefits for himself), (3) Intentional readiness for change is how employees are prepared so that energy is directed to the process of change.

This study is adapting scale Readiness for Organizational Change (Holt et al., 2007), because based on the above scale of Readiness for Organizational change can be used as a reference in advanced research, both in academia and industry as well as the organization and this measure is commonly used by RFC researchers for organization.

B. Previous Measurement Constructs

Based on adaptation of Readiness for Change conducted previously in Zulkarnain and Hadiyan (2014) The results of factor analysis for appropriateness dimensions obtained six-item with a value of MSA move from 0.823 to 0.899, and the value of the loading factor that moves from 0.679 until 0.845. On the dimension of change, there are five specific efficacy item with a value of MSA move from 0.569 to 0.827 and the value of the loading factor to move from 0.593 to 0.850. Furthermore, for the dimensions of the five-item management support that have moved from the MSA value of 0.747 to 0.793 and the value of the loading factor to move from 0.763 to 0.882. As for the personal dimension valance, obtained five item that has a value of MSA move

from 0,729 to 0.771, and the value of factor loading of 0.687 up to 0.810. This questionnaire consisted of 21 item with Alpha coefficients for 0928.

C. Measurements of this Study

Holt et al. (2007) expressed readiness for change as employee confidence in the ability of its self to change (self-efficacy), Appropriateness to make changes (appropriateness), support and commitment from management (management support), and changes confer any benefits to employees (personal benefits). In addition Holt et al. (2007) explains that employees who are ready for change will exhibit behaviors accept and support the plan changes.

Measurements on the readiness for change will be implemented using the adoption of an instrument that Holt et al. (2007) developed which consists of four dimensions: appropriateness, Management support, Change efficacy, and Personally Beneficial (personal valence) using techniques Confirmatory Factor Analysis (CFA) through Lisrell software version 8.7, in contrast to a previous study using Cronbach Alpha as analysis tools.

Table 1. Readiness for Change Dimensions (Holt et al., 2007)

Dimension	Item Numbers		Sum
	Favorable	Unfavorable	
<i>Appropriateness</i>	2, 3, 4, 5,10, 16, 25, 28, 32,38, 40	36,44	13
<i>Management support</i>	7,18, 24, 29, 31,33,39, 43	11,34	10
<i>Change efficacy</i>	13, 14,22, 27, 37, 42	19,20,30	9
<i>Personally Beneficial (personal valence)</i>	1, 8, 9, 12, 21, 26	15,17,41	9
Total	41		

II. METHODS

A. Subject Profiles

Table 2. Subject Profiles

Information	Profile	Frequency	Percentage
Sex	Woman	24	6,0%
	Man	376	94,0%
Education	Master's Degree	1	0,3%
	Bachelor's Degree (Engineering)	4	1,0%
	Bachelor's Degree (Non-Engineering)	21	5,3%
	Diploma's Degree (Engineering)	2	0,5%
	Diploma's Degree (Non-Engineering)	10	2,5%
	High School	281	70,3%
	Middle School	58	14,5%
	Elementary School	23	5,8%

B. Process Adaptation

Measuring instrument will be used is Holt et al. (2007) questionnaire, and the stage of adaptation (Beaton et al., 2000) is:

a. Phase I: Translating early (initial translation)

At this stage, the initial diversion from the original language to the target language. At least two people early translators with a different background is recommended to do so. In the process, the translator may consider the ambiguity of each word. Every translator produces a written report which consists of items, answer choices, and user charging, or he results of discussions both translators.

In this study, the initial translation is done with the help of a linguist who is a lecturer in scientific English (TOEFL), whereas content experts is a graduate of the Master of Psychology, which has now been in the field of Psychology of Industrial & Organizational.

b. Phase II: mensistesis the translation (synthesis of the translations)

The results of the second translator in the first phase, synthesized into one common translation (combined). The aim is to choose the most appropriate meaning to the initial scale.

c. Phase III: Translation back (back translation)

This stage is carried out by two people to do the translation of the results of phase two in the original language without knowing the initial questionnaire. This is done as a process to check the validity and ensure that the translated version of the content as the version reflects the actual (original).

In this study, back translation performed by lecturers PIO (Psychology of Industrial & Organizational)

d. Stage IV: The expert committee (expert committee)

The expert committee consolidation of all the translated version to be developed into a prefinal version. (1) Equality semantics. (2) Equality idiomatic. (3) Equality experience. (4) Conceptual equivalence. Phase V: Testing of prefinal version (test of the prefinal version)

3. Validity Construct

Uses CFA (confirmatory factor analysis) method. The logic of CFA according in Umar (2012) is: (1) Concept or trait that defined operationally called factors, and measurement of these factors is done through an analysis of the items response, (2) Item only measure one factor alone, that means either items or subtests are unidimensional, (3) If correct (unidimensional) then of course there is no difference between the matrix Σ - S matrix or can be expressed by $\Sigma - S = 0$, (4) The statement made null hypothesis is then tested by chi square. If the results of chi-square is not significant ($P\text{-value} > 0.05$), the null hypothesis is accepted. This means that unidimensionalities of theory can be accepted that the item or sub-test instrument measures only one factor alone, (5) If the model fit, then the next step to test whether or not a significant item measure what is to be measured, using the t-value. If the t-value is not significant ($t > 1.96$), it is not a significant item in measuring what is to be measured, if necessary, such items in drop and vice versa, (6) If the results of the CFA are items load factor coefficient is negative, then the item should be dropped. The CFA analysis testing like this is done by using LISREL software 8.70.

III. RESULT

A. Validity Construct for every dimension:

a. Appropriateness

The result of CFA is conducted using one-factor model. Chi-square = 77.87, df = 60, P-value = 0.06030, and the value of RMSEA = 0.036. P-value > 0.05 so it can be stated that the model with one factor acceptable. It means all items only measure one factor: Appropriateness.

The author investigated if the item measured the factor to be measured significantly and determined if the item should be dropped or not; the testing was done by looking at t value of every factor loading coefficient, such as in table 1 below.

Table 3. factor loading of Appropriateness

Item	Coefficient	Error Standard	t-value	Significance
2	0.65	0.07	9.85	√
3	0.49	0.07	7.16	√
4	0.51	0.07	7.55	√
5	0.53	0.08	6.75	√
10	0.80	0.07	11.40	√
16	-0.05	0.07	-0.66	X
25	0.16	0.07	2.22	√
28	0.04	0.07	0.62	X
32	0.03	0.07	0.49	X
36*	0.19	0.07	2.73	√
38	0.09	0.07	1.22	X
40	0.11	0.07	1.54	X
44*	0.06	0.07	0.88	X

√ Significance ($t > 1.96$); X = non-Significance ($t < 1.96$); * = unfavorable item.

b. Management Support

The result of CFA is conducted using one-factor model. Chi-square = 48.29, df = 34, P-value = 0.05312, and the value of RMSEA = 0.042. P-value > 0.05 so it can be stated that the model with one factor acceptable. It means all items only measure one factor: Management Support.

The author investigated if the item measured the factor to be measured significantly and determined if the item should be dropped or not; the testing was done by looking at *t* value of every factor loading coefficient, such as in table 2 below

Table 4. factor loading of Management support

Item	Coefficient	Error Standard	t-value	Significance
7	0.16	0.07	2.24	√
11*	0.06	0.07	0.97	X
18	-0.15	0.07	-2.07	X
24	0.10	0.08	1.48	X
29	0.09	0.07	1.39	X
31	0.19	0.08	1.39	X
33	1.00	0.20	4.91	√
34*	0.33	0.09	3.62	√
39	0.18	0.07	2.45	√
43	0.08	0.07	1.14	X

√ = Significance ($t > 1.96$); X = non-Significance ($t < 1.96$); * = unfavorable item.

c. Change Efficacy

The result of CFA is conducted using one-factor model. Chi-square = 32.71, df = 26, P-value = 0.17074, and the value of RMSEA = 0.033. P-value > 0.05 so it can be stated that the model with one factor acceptable. It means all items only measure one factor: Change Efficacy.

The author investigated if the item measured the factor to be measured significantly and determined if the item should be dropped or not; the testing was done by looking at *t* value of every factor loading coefficient, such as in table 3 below.

Table 5. Change efficacy

Item	Coefficient	Error Standard	t-value	Significance
13	0.31	0.10	3.12	√
14	0.75	0.18	4.27	√
19*	-0.23	0.09	-2.49	X
20*	-0.20	0.09	-2.21	X
22	0.03	0.07	0.33	X
27	0.11	0.09	1.35	X
30*	0.23	0.09	2.51	√
37	0.33	0.09	-0.75	X
42	0.08	0.08	0.90	X

√ = Significance ($t > 1.96$); X = non-Significance ($t < 1.96$); * = unfavorable item.

d. Personally Beneficial (personal valence)

The result of CFA is conducted using one-factor model. Chi-square = 35.74, df = 26, P-value = 0.096590, and the value of RMSEA = 0.040. P-value > 0.05 so it can be stated that the model with one factor acceptable. It means all items only measure one factor: Personally Beneficial (personal valence)

The author investigated if the item measured the factor to be measured significantly and determined if the item should be dropped or not; the testing was done by looking at *t* value of every factor loading coefficient, such as in table 4 below.

Table 6. Item Personally Beneficial (personal valence)

Item	Coefficient	Error Standard	t-value	Significance
1	0.34	0.08	4.30	√
8	0.83	0.11	7.42	√
9	0.31	0.08	3.98	√
12	0.49	0.09	5.71	√
15*	0.09	0.08	1.21	X
17*	0.00	0.08	0.01	X
21	-0.04	0.08	-0.47	X
26	0.09	0.08	1.19	X
41*	0.06	0.08	0.82	X

√ = Significance ($t > 1.96$); X = non-Significance ($t < 1.96$); * = unfavorable item.

B. Adaptation of the measurement results

NO	DIMENSION	FACTOR	
		VALID	DROP
1	Appropriateness	2,3,4,5,10,25,36	6, 28, 32, 38, 40, 44
2	Management support	7, 33, 34, 39	11, 18, 24, 29, 31, 43
3	Change efficacy	13, 14, 30	19, 20, 21, 27, 37, 42
4	Personally Beneficial (personal valence)	1, 8, 9, 12	15, 17, 21, 26, 41
Total		18	23

Based on the table above, the total item valid measuring adapted Readiness for Change into 18 items. This is not in line with a previous study that found 21 valid items (Zulkarnain & Hadiyan, 2014).

IV. CONCLUSION

Based on the result of validity tes, it concluded that the items of scale is significantly unidimensional, which is only measure the construct dimension Organizational Readiness for Change. In addition the process of translation is also deemed a success, because there is no difference in the meaning of the results backtranslation with initial scale. Respondents could also be easily understood the statements contained.

Furthermore, the results of validity of the test could be used as a reference for any research using the same variables. In addition, this measure can also be used for companies or organizations to determine the level of employee's readiness of change.

PREFERENCES

- [1] [1] Barrett, D. J. (2002). Change communication: using strategic employee communication to facilitate major change. *Corporate Communications: An International Journal*. 7. 219-231.
- [2] [2] Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). *Guidelines for the process of cross-cultural adaptation of self-report measures*. Spine, 25(24), 3186–91.
- [3] [3] Bouckennooghe, D., Devos, G., & Broeck, D.V.H. (2009). Organizational Change Questionnaire-Climate of Change, Processes, and Readiness: Development of New Instrument. *Journal of Psychology*, 143, 6.
- [4] [4] Elving, W.J.L. (2005). The role of communication in organisational change. *An International Journal*. 10. 129-138.
- [5] [5] Hanpachern, C., Morgan, G.A., Griego, O.V., (1998). An Extension of the Theory of Margin: a Framework for assessing readiness for change. *Human Resources Development Quarterly*, 9(4), 339-350.
- [6] [6] Holt, D., Armenakis, F.S., & Harris, G. (2007). Readiness for Organizational Change The Systematic Development of a Scale. *The journal of applied behavioral science*, 43, 232-255.
- [7] [7] Mangundjaya, W.L.H. (2012): Are organizational commitment and employee engagement important in achieving individual readiness for change?. *Humanitas*, IX, 2.
- [8] [8] Proctor, D & Doukakis, F, 2003. *Managing organizations in era of change*. Florida: The Dryden Press.
- [9] [9] Saragih, E.H., Hutagaol, P., Pasaribu, B., & Djohar, S. (2013): Individual attributes of change readiness in Indonesian television companies experiencing corporate transformational change – a quantitative approach using structural equation modeling. *International Journal of Innovations in Business*, 2, 1.
- [10] [10] Umar, J. (2010). Bahan penelitian analisis faktor. Institute Assesmen Indonesia.
- [11] [11] Zulkarnain., Hadiyan, S., (2014): Peranan Komitmen Organisasi dan Employee Engagement terhadap Kesiapan Karyawan untuk Berubah. *Jurnal Psikologi* 41(1), 17 – 3.

DEVELOPING ACHIEVEMENT TESTS IN PHYSICS FOR CLASSROOM ASSESSMENT

Dhien Astrini¹, Kumaidi²

¹Student of Yogyakarta State University;

²Lecturer of Muhammadiyah University of Surakarta/Yogyakarta State University;
dhienastrini@gmail.com

Abstract— Achievement test is required in the process of learning in purpose to get information and to evaluate the effectiveness of educational program. Giving tests to the students are expected to improve the quality of education which relates to the success of the learners in the teaching learning process. This study is purposed to develop an achievement tests in physics for classroom assessment and to get the characteristics of achievement test in physics of the tenth grade students of high schools. The models of test developments Orinondo and Dallo-Antonio model is used in this study. The trial tests are done in three schools which are categorized high, middle, and low. The subject of this study consists of 202 respondents of the tenth grades students of high schools. A qualitative analysis is done by reviewing achievement test instruments which are done by nine reviewers. In another hand, quantitative analysis is done by using item response theory which is analyzed by BILOG-MG program. The reviewers analyses are conducted based on grating tests, test specifications, test instruments sheet, and 40 test items. The results of analysis by using BILOG-MG program, based on two parameters logistic model, show that about 35 from 40 items are categorized as fit items. The coefficient reliability of achievement test is 0,86. It can be concluded that the achievement test instruments fit with two parameters logistic model.

Keywords : *Achievement Tests, Physics, Classroom Assessment*

I. INTRODUCTION

Recently, educational quality improvement has been conducted sustainably to form competent graduates. The learning activities and assessment systems is part of efforts to improve the quality of education. The learning activities related to three things: the purpose of learning, the process of teaching-learning and the evaluation of learning outcomes. The purpose of learning to direct on the assessment which will be implemented out by teachers. The process of teaching-learning in the implementation of refers to the purpose of learning and evaluation of learning outcomes. Those are conducted to assess what extent the process of teaching-learning in accordance with the purpose of learning predefined.

Teachers can evaluate the learning outcomes by preparing a test in the teaching-learning process. The test is one of important aspect in the implementation of teaching-learning process. Implementation the process of teaching-learning performed by each teacher at the school in addition to the necessary precision in teaching are also necessary to assess the process of teaching-learning. One way that can be used to assess student learning outcomes is by given the test, so that teachers can determine the extent of the achievement level of competence of learners.

Assessment in the process of teaching and learning is done through three aspects: cognitive, affective, and psychomotor. Assessment on cognitive aspects done by assigning the test in the form of assignments and exams to measure thinking ability of learners. The mastery learning material learners in schools are expected consistent with the *Standar Kompetensi Lulusan (SKL)* in *Kurikulum Tingkat Satuan Pendidikan (KTSP)* that has been set by the government. Assessment of achievement of competencies conducted one through the test result of learning to the learners. The proper assessment can be realized if the provision of the tests used by teachers are properly conducted. Results achievement of good competency are expected to explain the ability of learners. The result is processed into information that can be used as a

material consideration in determining the learners master of the learning material that has been given.

Test is instrument or procedure that systematic to measure the sample behaviors by conducting a series of questions in a uniform way [8]. The purpose of conducting the test is to find out the achievement of learning or competency which has been achieved of students in a particular field. Result of the test is the information about characteristics of a person or a group. These characteristics can include a person's cognitive abilities or skills [2]. The test determines as good test if it is able to measure what should be measured, it can give reliable results. The tests result can be used by teachers to (a) decide an ability of learners relatively to other learners in the same test; (b) show the development of the ability of learners within a period of time in knowledge and specific skills; (c) show the evidence of understanding of a subject matter, knowledge or idea; and (d) predict the performance of learners in the future [1].

Educational achievement tests, or tests measuring specific knowledge and skills such as those measured by licensure, certification, or employment tests, we believe validity evidence based on test content is critical for validating test scores that are used to make inferences about examinees with respect to the targeted domain. For educational tests, defining the domain measured is typically accomplished by providing (a) detailed descriptions of the content areas and cognitive abilities the test is designed to measure, (b) test specifications that list the specific content "strands" (sub-areas), as well as the cognitive levels measured, and (c) specific content standards, curricular objectives, or abilities that are contained within the various content strands and cognitive levels. For achievement testing in elementary, middle, and secondary schools, the content and cognitive elements of the test specifications are typically drawn from curriculum frameworks that guide instruction [12].

Problems related to achievement test on learning of physics, such as the item test for end classroom exam that created by the teachers is still found items are presented with the image, but not yet sufficient. In fact there are some images that have not been clear or not according to the physics problem on items. It should be conducted analysis of qualitative and quantitative to know the characteristics of items to determine good item. Physics is science about physical world that the most of fundamental, learn the basic principles of the universe. The main purposes of physics is looking for basic from various natural phenomena and using the related laws to develop theories that can predict in the future experiments [5].

The examination is techniques to obtain necessary information for evaluation purposes. The test are instruments that uses to obtain such information. Test provides teachers with information that can help teacher in improving instruction. Teacher also gives the learners about the information that can help them to understand whether they are good or not [9]. The step of development test (based on the construct and content validity) becomes important because it is used to find out the characteristics of the test [7]. Finally, achievement test required in the learning process to gather information and evaluate the effectiveness of educational programs. Thus, a test becomes one of the means that can assist all of the teachers, learners and parents to obtain the required information to improve the quality of education related the success of learners.

Developing achievement tests preceded by formulating content of physics and preparing grating tests. Formulation matter of physics on the purpose of measurement, so that items produced precisely measure the ability of learners to be measured on specific contents. A measurement model that use in the developing achievement test in physics is IRT models. IRT models are statistical models of the relationship between a person's value to the construct being measured and the probability of choosing each response test participants at any items to measure the construct. IRT models can be used to evaluate how the information items on a variety of specific scores and to estimate the value of a person IRT [6]. The purpose of this study was to developing achievement test in physics for classroom assessment and obtain characteristics a learning achievement test on physics.

II. METHOD

This study uses research and development test. The following test steps in order to build a systematic procedure are; (a) determine the purpose. The purpose of measurements were performed to produce precise items in the measurement; (b) prepare a table of specifications; table of specifications (blue-print) was based on the contents of the specified on formulation and then determine the percentage of the amount of items produced; (c) selecting an appropriate

items format; items format used in this study is multiple-choice; (d) writing test items; writing test items performed by grating tests; (e) editing the test items; before testing was conducted editing items should be done to avoid errors in the concept of the content; (f) the test trials; test items that have been edited furthermore tested on a limited group; (g) preparing the final form of the test; the final form of the test includes items that meet the criteria of a good items characteristics (h) establish the validity of the test; determining the validity of the tests conducted based on the contents (i) establish the reliability test; in developing the tests required constancy of a test to be used so that is indispensable high reliability coefficient (j) interprets the test; the interpretation of tests is the final step in the model development being done. The step in building tests are presented in the figure as followed.

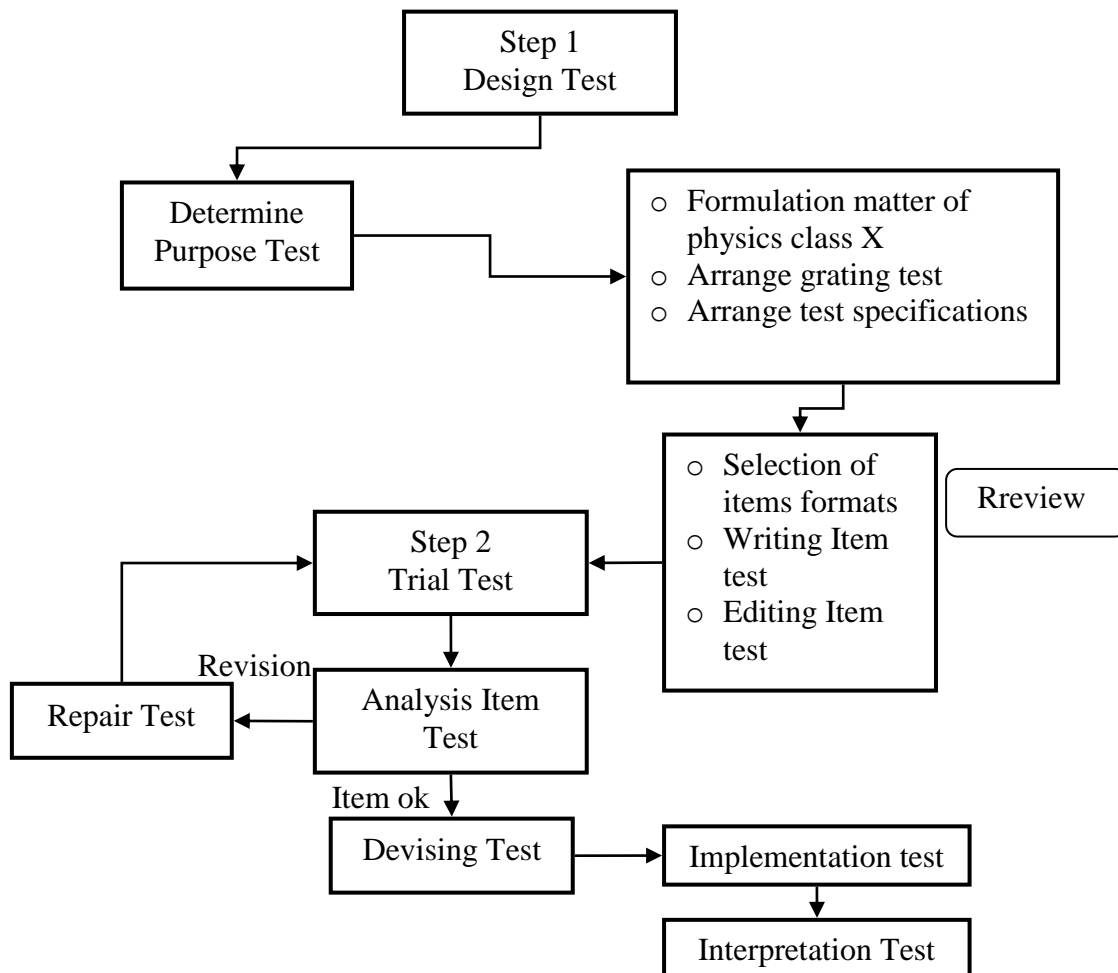


Figure 1. Research Design

The trial test conducted in three schools that have high category, medium category and low category. The trial subject of the tenth grade students were 202 respondents. Qualitative data analysis is done based on the review of the paper instrument achievement test. Qualitative data analysis conducted of review achievement test by nine reviewers. Determining the quality of instruments is done through content validity and empirically analysis. Empirically analysis to calibrate the test items will be obtained item fit to the model. Fit items with the parameters logistic model based on the analysis of items on parameters logistic model one, two and three parameters logistic. Test items are categorized fit to the model if proportion of correct or incorrect exceeds 0.05 ($\text{Prop} > 0,05$). Fit statistics on program requirements Bilog-MG, the cases in calibration samples are sorted into successive intervals of the latent continuum according to estimate of ability rescaled to mean 0 and standard deviation 1 [11]. If the data fit with the model means, it can be used to construct of the test.

III. RESULTS AND DISCUSSION

Developing achievement tests in physics conducted in accordance with the procedures of the study. Implementation starting with the first step by designing test begins determine the purpose test to measure the achievement of learners in the matter that has been taught by teachers. This relates to formulate material physics tenth grade used as ingredients test administration. Matter of physics formulated for use as an ingredient provision of the test consists of several sub-content with due regard syllabus of learning. The next step is to develop a grating about the based on the *Standar Kompetensi (SK)*, *Kompetensi Dasar (KD)*, and standards developed into several standards items in accordance with the range of material of physics. Gratings tests produced subsequently formed into a table of specifications (Blue print) to determine the scope of cognitive aspects to measure the percentage of the number of item cognitive aspects. Reviewing is done by providing suggestions and input to the item that has been written. Reviewing conducted by nine reviewer consists of measurement experts, education experts of physics, and teacher of physics at SMA Negeri Yogyakarta. Reviewing is conducted theoretically through the paper instrument ratings achievement tests. Validation produces some recommendations and input from the reviewer to improve item that was not relevant in terms of language, content and accuracy of answers. The results obtained as much as 33 item meets the appropriate criteria based on the contents (matter). This is an example paper review of test instruments at number 24 is presented as follows.

Table 1. Test Instruments Sheet in Item Number 24

Butir Tes	
<p>24. Sebuah bola berongga terbuat dari besi (koefisien muai panjang $1,2 \times 10^{-5}/^{\circ}\text{C}$) pada suhu 30°C jari-jarinya 1 m. Jika bola tersebut dipanaskan sampai 90°C, maka pertambahan luas permukaan bola adalah sebesar...m^2.</p> <p>A. $9.60 \pi \times 10^{-3}$ B. $5.76 \pi \times 10^{-3}$ C. $2.88 \pi \times 10^{-3}$ D. $2.40 \pi \times 10^{-3}$ E. $1.44 \pi \times 10^{-3}$</p>	
<p>Silahkan berikan penilaian (\checkmark) dengan nilai 1-4, sejauh mana butir fisika sesuai dengan standar kompetensi dan indikator soal yang disajikan berikut:</p> <div style="display: flex; justify-content: space-around; align-items: flex-end;"> <div style="text-align: center;">1 <input type="checkbox"/></div> <div style="text-align: center;">2 <input type="checkbox"/></div> <div style="text-align: center;">3 <input type="checkbox"/></div> <div style="text-align: center;">4 <input type="checkbox"/></div> </div>	<p>Standar Kompetensi : Peserta didik dapat menerapkan konsep kalor dan prinsip konservasi energi pada berbagai perubahan energi</p> <p>Indikator soal: Diketahui koefisien muai panjang, perubahan suhu, dan jari-jari benda, peserta didik dapat menganalisis pertambahan luas benda</p>
<p>Beri tanda (\checkmark) pada kotak di bawah. Jika butir sesuai dengan indikator (dinilai 3-4), maka silahkan beri tanggapan "Tidak" Jika butir tersebut tidak sesuai dengan indikator (dinilai 1-2), maka silahkan beri tanggapan "Ya"</p> <p style="text-align: center;">Tidak <input type="checkbox"/> Ya <input type="checkbox"/></p>	
<p>Saran perbaikan:</p>	

The second step is the trial test imposed on students of tenth grade were which 202 respondents. The data has been collected in validating empirically using the data of dichotomy through bilog-mg program. The analysis is done in stages which started by using model of 1 parameter Logistic, 2 parameter logistic and 3 parameter logistic. Analysis performed using three models to select the fit model of test items. The model produces items model fit at most it will be chosen as a model for the analysis of items.

Test items are categorized fit to the model if proportion of correct or incorrect exceeds 0.05 [11]. The Instruments achievement tests in physics had analyzed 40 items. Number the results of analyzes empirical data showed that the model which produced item fit with most models are two parameter logistic models. The analysis showed that as many as 36 items fit with models. The results of analyzes the item in the second phase of Bilog-MG program are 1 item uncalibrated that is item number 2, so that the number of items from which as many as 35 items. This is due to the phase 1 showed that the coefficient point biserial <0.3 so that the slope of the test items is very small.

The coefficient of reliability is based on item response theory through the results of the phase 3 Bilog-MG was obtained for 0.86. The reliability refers to the stability or consistency of the test scores. The reliability test of 0.86 provides information that measuring learning achievement tests in physics with consistent results. The higher reliability of a test is the more sample tests that can provide the expected information [10]. With a reliable measuring instrument, the measurement results will be the same information even if different examiners, correctors different or different items but measure the same thing and have the same characteristics of the item [4].

The results of data analysis using 2 models parameter logistic to show different power and the level of difficulty of point in phase 2. The values of b_i vary (typically) from about -2.0 to +2.0. Values of item difficulty near -2.0 correspond to items that are very easy, and values of b near 2.0 correspond to items that are very difficult, for the group of examinees and item discrimination parameters is between 0-2 [3]. The mean discrimination power parameters of item tests is 0.938 with a standard deviation of 0.618 and the mean item difficulty of tests is -0.622 with a standard deviation of 1.094.

Based on the analysis, item discrimination parameters is within the interval 0.253 to 1.98 whereas items difficulty is within the interval -1.75 to 0.878. This means that as many as 38 items that meet the criteria for item discrimination parameters is good and as many as 36 item that meet the criteria of a good difficulty of items. Seen in the figure below items that have a high degree of difficulty is in item number 35 with the items of difficulty of 0.878.

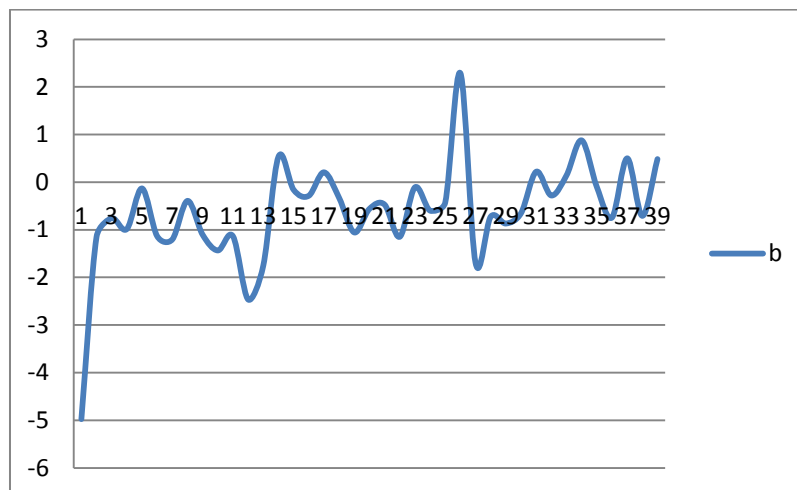


Figure 2. Item Difficulty

Based on the results obtained by analysis of two parameter logistics model so that it can be made the image of Item Characteristic Curve (ICC) as in Figure 3. In the test consists of six items that represent the tests. The shows that there are 3 items test that have a near level of difficulty and discrimination power parameters. Item 1 ($a=1.081$; $b=-0.984$), item 2 ($a=1.176$; $b=-0.28$) and item 3 ($a=1.512$; $b=-0.87$). It can be represented by three other items.

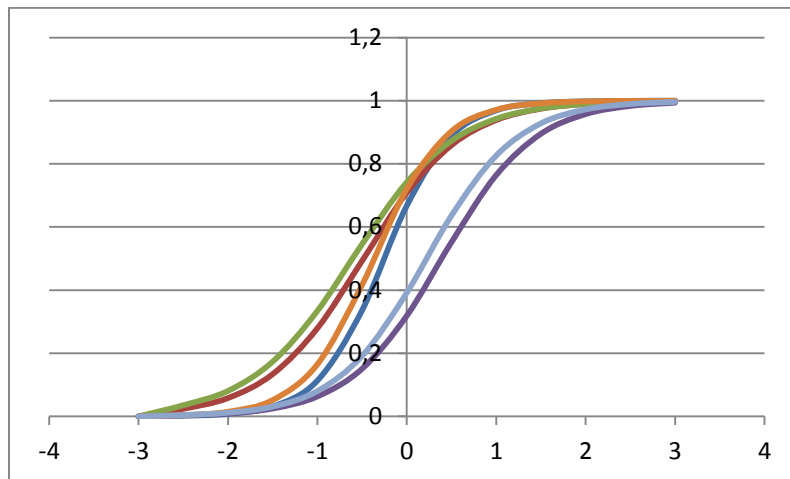


Figure 3. FIGURE 3. The Item Characteristic Curves of 6 items

Using of two parameters logistic model, item difficulty and discrimination power parameters can be estimated. Based on these result, Value of Information Function (VIF) can be estimated. The estimation are summed then, the Standard Error Measurement (SEM) can also be estimated using VIF. Correlation VIF and SEM results presented in FIGURE 4.

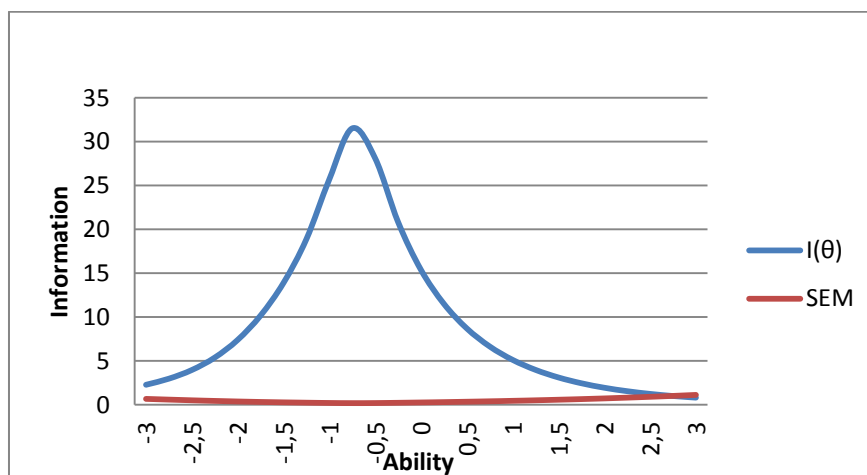


Figure 4. VIF and SEM

The value of test information function which is the total of the value of item information function. The information provided by a test at θ is simply the sum of the item information functions at θ [3]. In FIGURE 4. shows that the maximum value is 31.5 on a scale ability equals to -0.75. The amount of information provided by a test at θ is inversely related to the precision with which ability is estimated at that point where $SE(\hat{\theta})$ is called the standard error of estimation [3].

IV. CONCLUSION

The study successfully prepare grating test, table of test specifications (blue-print), test instruments sheet and the developing achievement tests in physics that form of multiple-choice. The test was developed consisting of 40 items. The analysis results obtained 35 test items fit with two parameter logistics model views based on the mean discrimination power parameters of item tests is 0.938 with a standard deviation of 0.618 and the mean item difficulty is -0.622 with a standard deviation of 1.094. The reliability values of 0.86 tests provide consistent results in providing information measuring learning achievement tests in physics. Finally, achievement tests in physics that have been developed can be used for end classroom exam as part of classroom assessment.

REFERENCES

- [1] Bambang Sumintono & Wahyu Widhiarso. (2015). *Aplikasi Pemodelan RASCH pada Assessment Pendidikan*. Cimahi: Trim Komunikata.
- [2] Djemari Mardapi. (2012). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Litera.
- [3] Hambleton, R. L., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.
- [4] Heri Retnawati. (2016). *Validitas, Reliabilitas & Karakteristik Butir (Panduan untuk Peneliti, Mahasiswa, Dan Psikometrian)*. Yogyakarta: Nuha Medika.
- [5] J. Serway. (2010). *Fisika untuk sains & teknik edisi 6*. Jakarta: Salemba Teknika.
- [6] Jakob B. B., & Chih-Hung, C et al. (2007). *Developing tailored instruments: item banking and Computerized Adaptive Assessment*. *Qual Life Res* Vol. 16: 95–108.
- [7] Lissitz W, Robert & Samuelsen Karen. (2007). *Further Clarification Regarding Validity and Education*. *Educational Researcher*, Vol. 36, No. 8, 482–484.
- [8] M. David Miller, Robert L., Linn & Norman E., Gronlund. (2009). *Measurement and Assessment in Teaching (10 th ed)*. Englewood Cliffs, NJ. : Prentice-Hall.
- [9] Oriondo, L.L.and Dallo-Antonio, E.M. (1998). *Evaluation Educational Outcomes*. Manila: Rex Printing Compagny, inc.
- [10] Reynolds, C. R., Livingston, R. B., & Willson, V. (2009). *Measurement and assessment in education (2nd ed.)*. USA: Pearson.
- [11] Robert J. Mislevy, R. Darrell Bock. (1990). *BILOG 3 Item Analysis and Test Scoring with Binary Logistic Models second edition*. USA: Scientific Software, Inc.
- [12] Sireci, S & Faulkner-Bond, M. (2014). *Validity evidence based on test content: Psicothema*. Vol. 26, No. 1, 100-107.

The Development of Evaluation Model Education Life Skill Program Out of School Education

Edi Subarkah

STKIP Catur Sakti, Yogyakarta
email : edisubarkah1960@yahoo.com

Abstract - Evaluation Model Education Life Skill Program Out of School Education is taken from some sources of theories and empirical experience in the field. There are some models which are developed by the experts and used in evaluating learning program. Kirkpatrick, with the training program in human and resource development field (HRD) offers evaluation model named Kirkpatrick Training Evaluation Model. Besides that, the most well known model applied by the evaluator is CIPP Evaluation Model Concept. This model is completed by one component of O, the abbreviation of Outcome (S). Therefore now the model is known as CIPPO. To strengthen the program evaluation model in this paper, the theories are taken from the elements program evaluated by Out of School Education from Djuju Sudjana. The Kirkpatrick model has some strengths. They are 1) more comprehensive because it covers hard skills and soft skills; 2) the evaluation object is not only the result of study but also the process, output and outcomes; 3) easier to be applied for the class level because it doesn't too much involve other people in evaluation process. Kirkpatrick model has some limitations. Such as 1) it gives less attention to input because the limitation of the output in learning process is also affected by the input; 2) In measuring the impact, it is difficult to measure the parameter and it is also out of the teacher and the school's scope. CIPP model has some strengths. They are more comprehensive, because the evaluation object is not only to the result but also the contexts, input, process and result. The limitation of CIPP is the application of this model in learning program in the class has the low degree of implication without modification. The evaluation of out of school program by Djudju Sudjana has the strength in input, education process through learning, output and outcome. The measurement of outcome is generally hard to describe in short time but it can be observed minimally through knowledge, attitude, and the action of the students.

Keywords : education, life skill, out of school

I. INTRODUCTION

The implementation of education national system includes sub system of non formal education, often gets sharp sighted from the society and from the other institution since the policy and the implementation is often changed without accurately supported data. One of the factor is the lack of evaluation which is done regularly and continuously. The evaluation should be done continuously to the input, process, result and impact of non formal education.

There are many evaluation model which is being done in evaluation. Some of them are theoretical design which is proposed by the expert, some evaluation field experiences which is developed, some are concepts, guidance, and clue to do program evaluation. Kirkpatrick which is one of expert in training evaluation program and human resource development not only offers evaluation model named Kirkpatrick Evaluation Model but also shows the other model which can be used as choice in evaluating a program. Suharsimi Arikunto and Cepi Safrudin Abdul Japar (2008: 4) said, "There are some model evaluation which is known as the determiner of model evaluation program. They are Stufflebeam, Metfessel, Michael Seiven, Stake and Claser, Kaufman and Thomas divide into eight". The 8 kinds of evaluation model are : 1) Goal Oriented Evaluation Model which is developed by Tyler. 2) Goal Free Oriented Evaluation Model which is improved by Scriven. 3) Formative Sumatif Evaluation Model which is proposed by Michael Scriven. 4) Countenance Evaluation Model which is developed by Stake. 5) Responsive

Evaluation Model which is improved by Stake. 6) CSE-UCLA Evaluation Model which emphasize on when the evaluation is conducted. 7) CIPP Evaluation Model which is proposed by Stufflebeam. 8) Discrepancy Model which is developed by Provus.

From those some models above, this essay will only discuss about evaluation model by Donald L. Kirkpatrick and CIPP which are popular and used by many people as strategy or working guidance in the implementation of program evaluation. Besides two statements above, Djuju Sudjana (2006 : 89 – 99) about out of school evaluation program gives contribution the model of evaluation program. With the consideration about the strength and weaknesses of those 3 models, the writer modifies to produce the suitable model to evaluate education program.

Out of school education is done to give life skill to the students. Life skill is a skill to adapt and behave positively which allows individual to react effectively to face the needs and daily challenges. The policy of out of school education will have a mean when it gives life skill to the students in the implementation and the students can get prosperity in their life. The hypothesis in this paper is What are the strengths and weaknesses of evaluation model in non formal education program?

II. THEORETICAL REVIEW

A. *Life skill education in out of school education*

Definition of life skill education.

The term of life skill education is meant by the skill which is owned by someone to be brave to face the life problem and the normal life without feeling stressed and then proactively and creatively try to find the solution and finally can solve the problem.

Brolin, D. E (1989) said that life skill constitute a continuum of knowledge and aptitude that are necessary for a person to function affectively and to avoid interruptions of employment experience. In this case, life skill education is as the continuum of knowledge and ability which is needed by someone to be independent in life. The other statement also said that life skill is a skill which must be possessed by someone to be happy in life. Based on the statement above, we can say that the life skill education is a knowledge and ability which is done by someone to get the better life.

The statement from Santoso S. Hamijoyo (2002), the idea of life skill education has been begun in 1949 from UNESCO through the concept of functional literacy. The main idea from that concept is that the ability of reading, writing, counting can give benefit for the someone to overcome 3 miseries. They are ignorance, ill-health and poverty. The functional literacy concept is the obligation for all the nation to supply, prepare, fix, increase and guarantee the quality about the implementation of life skill education especially the important life skill therefore the society can get the benefit and be spread evenly.

Slamet PH, (2002: 4) defines life skill as the skill besides the academic oriented skill. Based on that statement we can say that life skill besides to the academic oriented also the activity of working, independent effort and joint effort. Ditjen Diklusepa (2003: 6) shares the essence “ Life skill oriented education in out of school education is an effort to increase knowledge, attitude and skill which make the students to life independently”. The implementation of life school education in out of school field is based on five (5) education pillars. They are learning to know, learning to do, learning to be, and learning to life together. Based on that statement we can say that the life skill oriented education is a program which can study to improve knowledge and skill which is had, benefitted the specific knowledge and skill to increase the life quality and also help the needed people.

According to Anwar (2004: 20), the life skill education program is an education which can give practical and useful skill related to the needs of the job market, the business opportunity and economic potential or industry in the society. Life skill refers to many abilities which is needed by someone to get life successfully, happily and have a value in the society. Life skill is the ability to have communication effectively, to improve cooperation, to do the role as responsible citizens, to have preparation and skill to work and to have character and ethic to work. Based on that statement we can say that life skill is expected to help someone to have self esteem and confidence to earn money in the context of opportunity in the environment.

Ditjen PLSP (2004) gives definition that life skill education is the ability or skill to adapt and behave positively which allows someone to face challenges in his life effectively. From the

statement we can say that the students are expected to be able to increase the knowledge and skill which is had, to use the knowledge and skill to increase life quality and also to help the needed people.

According to Direktorat Kursus dan Kelembagaan (2010: 2), it is said that The life skill education which is done by training and institutional directory to give chance to study to get knowledge, skill and develop creative mental attitude. Innovative, responsible and dare to face the risk (professional mental attitude) in entrepreneurship in the way to increase the life quality.

Based on the statement above we can say that life skill education is wider than the skill to work since the working or unemployment person needs life skill to increase the life quality.

Life skill based on WHO (1977), classsified into five (5) groups. They are self awareness or personal skills, social skill, thinking skill, academic skill, and vocational skill. This definition can be said as academic, vocational, personal and social skill. While from United States Labor Office which is said by H Moh Alifuddin (2011: 43) stated that life skill is the daily skill needed by someone to be success in his/ her life. Eventhough there are different statement to define life skill, but the essential is that life skill is ability, skill and capability needed by someone to face and run the real life. Based on the above statement it is said that life skill gives conclusion to refer to many ability and sincerity needed by someone to go through the real life succesfully and happily.

Based on several experts' statements and explanation above so it can be said that life skill education is an ability, skill and capability needed by someone to face and run real life to work independently, creatively, innovatively and responsibility.

B. Model Evaluation Program

There are so many evaluation model improved by the experts which can be used in evaluating learning program. They are :

1. Evaluation model Donald L. Kirkpatrick

Evaluation model which is developed by Kirpatrick has done some revisions, the latest version is in 1998 in Kirkpatrick's book entitled Evaluation Training Program: The Four Leves (Eko Putro Widoyoko, 2009: 173). The Kirk Patrick's four levels evaluation model now becomes one of the references and standard for many big companies in training program for human and resources development.

The model improved by Kirkptrick is known as Evaluating Training Program: the four levels or Kirkpatrick's evaluation model. The evaluation to the training program covers 4 evaluation levels such as reaction, learning, behaviour and result.

Evaluating reaction is to measure the participants' satisfaction. To make the training effective, so the participants of the training need to react positively and favorably. According to Donald L. Kirkpatrick (1998: 19), reaction is meant as the word reaction implies, evaluation on the this level measure how those who participate in the program react to it. I call it a measure of customer satisfaction. Eevaluation in this level to measure how people participate in the program.

Learning can be defined as the extend to which participants change attitudes, improve knowledge and or increase skill as the result of attending the program. Therefore to measure the effectiveness of training program, the 3 aspects (attitude, knowledge and skill) need to be measured. Without the changing of attitude, the improvement of knowledge and the skill development of the training participants, so the program can be said to fail.

It is said by Donald L. Kirkpatrick (1998: 2) that behaviour can be defined as excent to which change in behaviour has accured because the participants attended the training program. The behavioural evaluation is different with the evaluation of attitude. The attitude evaluation in learning evaluation is focused on the changing of attitude which happen during the training so it is internal factor. While the valuation of behaviour is focused on the changing of behaviour after the participant back to work. Will the result of the learning be implemented after the participants back to work? So the behaviour evaluation is also called as external avaluation. What kinds of behaviour changing which happen in the working place after the participants join the training program? It should be given mark wether the participants feel happy after joining the training and back to work? How can the participants transfer knowledge, attitude and skill which is gained during the training and implement it in the working place?. Since the mark is the changing of attitude after back to work so the evaluation in this behaviour evaluation can be stated as outcome evaluation. It is more complex and harder if compared with reaction evaluation and

learning evaluation. Based on Kirkpatrick (Eko Putro Widoyoko, 1988: 49) that behaviour evaluation can be done with comparing the behaviour of the control group with the behaviour of the participants of the training, or comparing the behaviour before and after joining the training or conducting a survey or interview with the coach, higher authority or lower authority of the participants after back to work.

The evaluation is focused on the final result which happen because the participants have already attended a program. Donald L. Kirkpatrick (1998: 23) said that the result can be defined as the result that occurred because the participant attended the program. Result can be defined as the final result which has been reached because the participants attend the program. On the other part is also said that the final result can include increased production, improved quality, decreased cost, reduced frequency and or severity of accidents, increased sales, reduced turnover and higher profits.

The model program evaluation Donald L. Kirkpatrick is very useful to evaluate life skill program for courses and out of school education since this program consists of knowledge, attitude and skill or psychomotor aspects. The other factor to strengthen the use of this model is that the participants are guaranteed to get a job or work by themselves therefore it's possible to monitor and evaluate how far does the trainee want and able to apply the changing of attitude, improving knowledge or the betterment of skill which gained in the training the job. In the Kirkpatrick's term of behaviour or outcome can be marked with the cooperation of friends or employee leader which has attended the training. If the participants of the training is the new employee of the organization so the impact evaluation is also through friends of the leader of the organization. To evaluate the impact of training in the organization such as the improvement of production, the reduction of cost, the increase of benefit and etc still can be evaluated. The training activity in out of school education is usually focused in the aspect of vocational education.

The model evaluation from Donald L. Kirkpatrick to the training program covers 4 levels of evaluation such as reaction, learning, behaviour and result.

The strengths and weaknesses of evaluation model Donald L. Kirkpatrick

Compared to the other model, Kirkpatrick model has some strengths such as They are 1) more comprehensive because it covers hard skills and soft skills; 2) the evaluation object is not only the result of study but also the process, output and outcomes; 3) easier to be applied for the class level because it doesn't too much involve other people in evaluation process.

While the limitations of the Kirkpatrick model are 1) it gives less attention to input because the limitation of the output in learning process is also affected by the input; 2) In measuring the impact, it is difficult to measure the parameter and it is also out of the teacher and the school's scope.

To overcome the limitation in no 2, the out of school education has already make cooperation with the user of the graduate participants in the beginning or before the training is held. There has been an agreement to use the participants which has finished to work in the company if the participants don't choose to work by themselves.

2. CIPP Evaluation Model

This evaluation model is the model that has been known and conducted by the evaluator. The concept of CIPP (Context, Input, Process and Product) is firstly offered by Stufflebeam in 1965 as his result to evaluate ESEA (The Elementary and Secondary Educational Act). Besides that, Stufflebeam also as the expert who proposes a decision oriented evaluation approach structured to help the administrator to make decision. The concept is offered by Stufflebeam with the view that evaluation is used to fix not to prove. Moreover, Madaus & Stufflebeam (1993: 118) said that the CIPP approach is based on the view that the most important purpose of evaluation is not to prove but to improve. CIPP evaluation model can be implemented in many field like education, management, company and etc and also in many levels for project or institution. According to Djuju Sudjanan (2006: 54) this model identify 4 types of evaluation program which related to 4 types of decision in planning a program. Evaluation context provides the data about the decision and the planning of the program, input evaluation provides alternative decision to control the program, and product evaluation to provide the decision about result and recycle of the program.

CIPP is the abbreviation of Context evaluation, Input evaluation, Process evaluation and Product evaluation. Those four words or CIPP is the target of the evaluation. Based on Suharsimi

& Cepi Safruddin Abdul Japar (2008: 46) this model is completed by the component O, the abbreviation of outcome (s) so the model becomes CIPPO. The CIPP model only measures output (product), while CIPPO model measures to the implementation of the product. So if the product stops only on the graduation, while outcome (s) until how is the participation of the graduation in the society or the next education. In the product of the factory is not only rely on the product quality but also to the satisfaction of the user or consumer.

The strengths of CIPP evaluation model are more comprehensive, because the evaluation object is not only to the result but also the contexts, input, process and result. The limitation of CIPP is the application of this model in learning program in the class has the low degree of implication without modification. Since to measure context, input and result need long time, much money and engage many people.

3. Djuju Sudjana

There are some learning process components such as environmental input, through nature environment, social and institution. The nature environments which should be done in out of school education are biological nature environment, non biological nature environment and synthetic nature environment. Instrumental input consists of curriculum or learning program, education human resources, infrastructure and expense. Curriculum or learning and learning program covers the objective of the learning, methods or techniques and learning result evaluation tools, techniques and learning media and also and learning result evaluation tools. The objective of the learning related to the general learning objective and specific learning objective or it can be said while learning objective and final learning objective. The learning objective is used to fulfill the need of education and the need of study. Learning material consists of materials which is arranged systematically and prepared to be studied by the students based on their learning. Methods, techniques and learning media are used from the learning strategy to help the students to gain the learning objective. Evaluation tool is instrument such as test or examination to measure how far is the students' attitude after learning contrasted with before learning activity. Education as the element of education human resources which covers basic ability, academic, personal, social and professional. Education consists of tutor, learning guardian and trainer.

Learning infrastructure consists as learning location, building and learning equipment. Learning infrastructure needs to be evaluated about the supply, the quantity, the quality, the suitable with the learning, the improvement and the maintain. Out of school organization which is evaluated is the organization, the rule which is used, type and organization status, support capability, network and cooperation relationship with related outside stake holders. Thus the input infrastructure variables which are evaluated cover the objective of the program, curriculum, human and resources education, infrastructure, organization, management and defrayal.

Raw input is the students which consist learning citizen, training participant and teacher candidate. The students have internal characteristics and eksternal characteristics,. The evaluation to the raw input is to answer question about what characteristics which is the most supported or blocked the students to learn? and how is the influence to the process, result and outcome of learning?

The other input is the sources or support capability which allows the graduation to implement the learning result in his/ her life. The other input can be classified into business field, working and society activity. The other input needs to be learned by the students in learning process.

Learning process is educative interaction between the teacher's main input (guidance's attitude, students' subject) with the raw input that is student through the learning activity, guidance, information and training. The learning approach used is continuum approach or continuous and recycle which is started from pedagogy and andragogy or vise versa. Education is the science and art of teaching children. While andragogy is the science and art to help adult. So education can be started from pedagogy and then to andragogy or vise versa. The learning is done participatively which the teacher engage the students in the activity of identifying need, formulation of the learning objective, arranging learning program and evaluation process, result and learning outcome.

The output is the graduation from out of school education program. The graduation which is evaluated is quality and quantity of program graduation after experiencing learning process. Quantity is the sum of the graduation which is succeeded to finish the learning

process in the education program. Quality is the changing of students' behaviour or graduation covers the affective or attitude, aspiration, feeling and psychomotor.

Why outcome monitoring should be done? Based on the statement of Joseph Wholey, Harry P. Hatry, Kathryn E. Newcomer (1994: 99) to the question of why outcome monitoring should be done, the outcome to some is self evident: outcome monitoring keeps those who are responsible (and others who are interested) apprised of performance.

The influence (outcome) is the impact which experienced by the students or graduation after getting the support from the other input. This influence can be measured specially in 3 aspects, the first is the increasing of the prosperity in life with the indicator of possessing business or job, income, health, education and self appearance. The second is the learning effort from the others to the individual, group or community. The third is the participation in the social activity or the building of society like participation in thought, power or skill. Therefore out of school education will complete if the program and the evaluation regarded to all elements in education system. They are environment input, infrastructure input, raw input, process, family, other input and outcome.

III. CONCLUSION

The development of program evaluation model life skill education is improved by some experts which has strenghts and weaknesses. Kirkpatrick has strenghts. They are more comprehensive because it covers hard skills and soft skills, the evaluation object is not only the result of study but also the process, output and outcomes, easier to be applied for the class level because it doesn't too much involve other people in evaluation process. Kirkpatrick model has some limitations. Such as it gives less attention to input because the limitation of the output in learning process is also affected by the input; and in measuring the impact, it is difficult to measure the parameter and it is also out of the teacher and the school's scope.

The strengths of CIPP evaluation model are more comprehensive, because the evaluation object is not only to the result but also the contexts, input, process and result. The limitation of CIPP is the application of this model in learning program in the class has the low degree of implication without modification. Since to measure context, input and result need long time, much money and engage many people.

Djudju Sudjana has strenghts in out of school education. They are environment input, raw input, other input, process, output, and also outcome. The weakness is the reaction aspect hasn't been included to react the participants satisfaction. To make the training run effectively, so the training participants need to react positively.

REFERENCES

- [1] Anwar. (2004). Pendidikan kecakapan hidup (life skill education). Bandung: Alfabeta.
- [2] Brolin, D. E. (1998). Life-Ocentered carrer education: A Competency - Based Approach. Reston VS: The Council for Exceptional Children
- [3] Direktorat Pembinaan Kursus dan Kelembagaan. (2010). Kursus wirausaha desa daerah tertinggal, pedoman untuk lembaga-lembaga PNF daerah tertinggal. Jakarta: Kementrian Pendidikan Nasional.
- [4] Djuju Sudjana. (2006). Evaluasi program pendidikan luar sekolah. Bandung: PT REMAJA ROSDAKARYA.
- [5] Eko Putro Warnono, S. (2009). Evaluasi program pembelajaran panduan praktis bagi pendidik dan calon pendidik. Yogyakarta: Pustaka Pelajar.
- [6] Joseph S. Wholey & Harry P. & Kathryn E. Newcomer. (1994). Handbook of Practical Program Evaluation. San Francisco: Berrett Publishers.
- [7] Kirkpatrick, Donald L. (1998). Evaluating training program the four levels. San Francisco: Berrett Publishers.

Development of Performance Assessment in Guided Inquiry Learning to Improve Metacognitive Skills and Student's Achievement

Endah Handayani¹, Sunarmi², Murni Saptasari³

¹Department of Biology, Faculty of Mathematics and Science, State University of Malang

^{2,3}Department of Biology, Faculty of Mathematics and Science, State University of Malang

¹Email: endah.handayani97@gmail.com

Abstract— Metacognitive skills is one of the 21st century skills that have an important role in the success of learning. Based on the results of interviews with biology teachers in High School 1 Lawang note that there has been no assessment of metacognitive in there. The development of metacognitive skills assessment and student achievement can be done by using performance assessment. Therefore, the purpose of this research to develop and test the feasibility of performance assessment in guided inquiry learning to improve metacognitive skills and student achievement in High School 1 Lawang on *Kingdom Plantae* chapter. The model of development used ADDIE model consists of five stages. Metacognitive skills measured by student self-reflection on each section of chapter. Research and development was tested for class X-MIA 7 High School 1 Lawang consist of 24 students. Results of research and development performance assessment shows the average level of validity is 3.79 with the criteria of "Valid" and the average level of practicality is 3.65 with the criteria of "High". Metacognitive skills showed an increase of 6.1% in the last meeting, and student's achievement increase until 91.67% with the criteria of "Very High".

Keywords: *guided inquiry, metacognitive skills, performance assessment, student achievement*

I. INTRODUCTION

Metacognitive skills is one way of thinking that must have to confront 21st century is important for students to develop a positive mental related to how to learn, learning restrictions, and as a guide if it will fail^[1], as well as a central component in the process of self-regulating learning^[2]. During this time teachers only emphasized the success of student learning assessment of the product, while the process of learning, such as students' metacognitive skills, not much attention. Agree with the statement, most teachers do more emphasis on the assessment of learning outcomes, while the learning process even less attention tends to be ignored^[3]. In fact, the learning process will determine learning outcomes. Based on interviews on 12 and 13 November 2015 to biology teacher at High School 1 Lawang known that biology teachers are not yet clearly understand the definition of metacognitive skills and do not know the techniques of metacognitive assessment. The teacher stated that basically the student has the knowledge and awareness of their own cognitive processes (metacognitive knowledge) but can not regulate or control themselves against cognitive process itself (metacognitive skills).

Assessment of metacognitive skills is a form of appreciation for students' learning process. But that does not mean learning outcomes assessment is not necessary. The result of learning is the ability of the student after he received work experience^[4]. Learning outcomes can be used as a benchmark achievement of learning objectives that can easily be observed teachers. Metacognitive skills assessment and student learning outcomes can be performed using performance assessment, which is a way of teaching and learning that involves simultaneously processes and products^[5]. Significant influence on the performance assessment of all the dimensions of metacognitive skills, clear performance assessment has a positive impact on students' metacognitive skills^[6]. Performance assessment which has no need to be developed to improve metacognitive skills and student learning outcomes. One form of performance assessment is a portfolio^[3,7,8,9,10].

Develop the performance assessment is a tool for assessment and further can be used as evaluation of learning, **it needs** methods and models that support learning. Goals, methods, and

evaluation is an integral and interconnected in a consistent, continuous, and harmonious, called the **Triple** Anchors Teaching^[5]. Therefore necessary methods and models appropriate learning to support the development of performance assessments aimed at improving metacognitive skills and student learning outcomes. Developing a classroom culture of inquiry and argumentation provides an excellent opportunity for teachers to engage students in scientific reasoning, decision making, and reflection-all important aspects of metacognition^[11].

One biological materials that require exploration of natural phenomena is the classification of plants (*Kingdom Plantae*). *Kingdom Plantae* material is not included in the material that is difficult to learn, compared with molecular biological material which can not be observed directly. But precisely because the material *Kingdom Plantae* can be easily observed in nature, so right to use guided inquiry learning model.

II. RESEARCH METHOD

Research and development is designed for learning material X grade Kingdom Plantae odd semester. The development model used in this research and development is a model *ADDIE* which consists of five stages: Analysis (analysis), Design (design) Develop (development), Implement (application), and Evaluate (ratings)^[12]. Each stage consists of several stages outlined in the research and development procedure. Design of experiments on research and development is done by three (3) phases, including (1) test the validity of the expert material, (2) test the validity of the expert assessment, (3) test the validity by field practitioners, (4) test the use of the small groups. After getting a valid product based on the results of validation from assessment expert, biology teachers as field practitioners, and material experts then tested on users' goals, namely the class X MIA-7 High School 1 Lawang consisting of 24 students as a test of the effectiveness and practicality.

III. RESULTS AND ANALYSIS

Based on the analysis of data validation by expert assessments showed that the eleventh developed performance assessment have criteria "valid", but need some revision. Based on data analysis by biology teacher showed that the performance assessment developed eight got a valid slightly revised criteria and performance assessment got three very valid criteria do not need revision. The average of validity can be seen in Table 1.

Based on the results of data analysis showed that the validation results Student Worksheet with the criteria "Valid Enough", so it needs to be revised in part. Average results of data analysis of student questionnaire responses and responses of teachers, can be seen in Table 2., indicates that the performance assessment have been developed that can be used.

Table 1. Average of Validity

Number	Validator	V α	Validity Criteria	Description
1.	Assessment Expert	3,88	Valid	little bit revised
2.	Biology Teacher	3,79	Valid	little bit revised
		$\Sigma V\alpha = 7,67$		
		$\overline{V\alpha} = 3,84$	Valid	little bit revised

V α : Validity value

Table 2. Average of Practicality

Number	Subject	IO	Practicality Criteria	Description
1.	Biology Teachers	3,65	High	Doesn't need to test again
2.	Students	3,31	High	Doesn't need to test again
		$\Sigma IO = 6,96$		
		$\overline{IO} = 3,48$	High	Doesn't need to test again

IO : Practicality value

The average of metacognitive skills showed that in first to the second meeting metacognitive skills decreased 4.6%, from the second to the fourth meeting metacognitive skills decreased 9.4%, while from the fourth to the fifth meeting of meeting metacognitive skills increased 6.1 %

(in third meeting didn't measure metakognitive skills). Completely value of metacognitive skills in each sub-chapter/meeting can be seen in Figure 1.

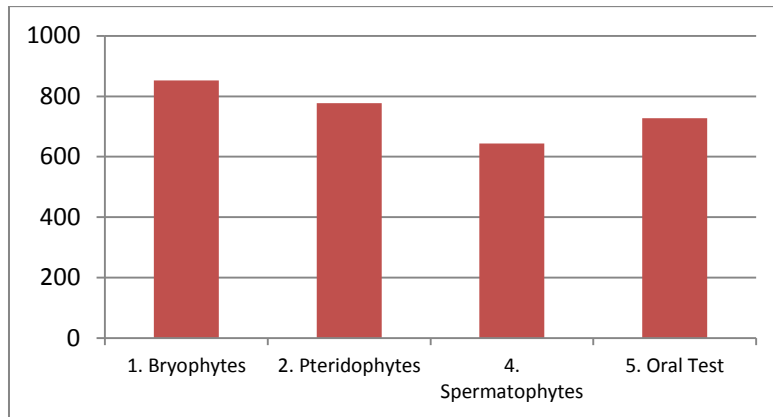


Figure 1. Value of Metacognitive Skills in Each Sub-Chapter of *Kingdom Plantae*

Metacognitive skills measured from self-reflection in this study consists of five indicators and have different percentages, can be seen in Table 3. The effectiveness of performance assessment in improving cognitive student learning outcomes measured from oral test, because it's final test on *Kingdom Plantae* chapters assessed using the rubric, so that the effectiveness of the performance assessment 91.67% with the criteria of "Very High", can be seen in Figure 2.

Table 3. Value of Metacognitive Skills in Each Indicators

Indicators of Metacognitive Skills	Number of Value	Number of Value Percentage (%)
A. Reveals work/strategies to improve the processes and learning outcomes	804	26,8
B. Reveals the advantages and disadvantages of the strategies used in gaining an understanding	640	21,3
C. Clearly Determining the target of learning outcomes	306	10,2
D. Reveal material that has been understood or not understood	644	21,5
E. Planning and organizing next learning activities	606	20,2
Total	3000	100

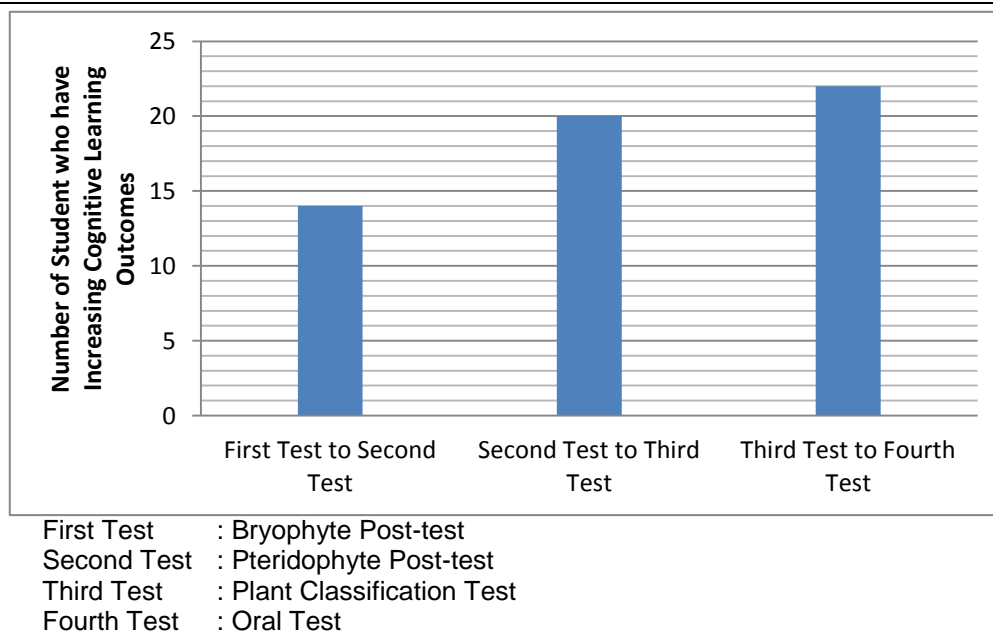


Figure 2. Effectiveness of The Performance Assessment

The average of psychomotor learning outcomes 100% above the KKM, with a mean value of 86.4. The average of affective learning outcomes of students 100% above the KKM, with a mean value of 84.1, can be seen in Figure 3.

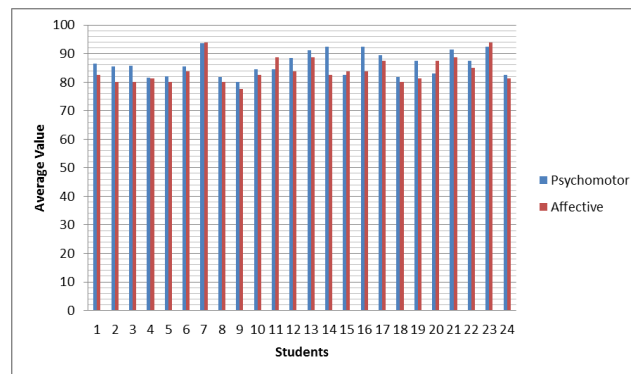


Figure 3. Average of Psychomotor Learning Outcomes

IV. DISCUSSION

In this study developed some performance assessment which aims to accommodate all the skills, processes, and products of learning from the students. Performance assessment that was developed not only to assess performance in the laboratory, but also the performance in the discussions, lab reports, and tasks in the Kingdom Plantae chapter. Hopefully, by the performance assessment can provide a complete report on the activities of student learning. In addition, the development of performance assessment is expected to improve metacognitive skills and ultimately improve student learning outcomes. To conduct the assessment required learning model is appropriate, for objectives, methods, and assessment are the three fundamental to learning, also known as the "Triple Anchors Teaching" ^[5].

Performance assessment developed in this study include observation sheet by teachers in practical activities, group discussions, class discussions, and presentations. Avoid subjectivits of teacher observation sheet is modified to be used as a self-assessment performance. In addition, well developed posters and portfolio assessment. All performance assessment validated by an expert assessment and field practitioners, and students' worksheet used during this study validated by subject matter experts. The assessment instrument must collect relevant data and information that are consistent, informative, reliable, and valid for all students ^[13]. Performance assessment developed has been validated by an expert assessment and field practitioners with a value of validation is 3.84 and further improved by the results of the validation, suggestions and comments from the validator.

Besides tested for validity, performance assessment developed also tested on a biology teacher and class X MIA-7 High School 1 Lawang which consists of 24 students. The results of these trials to get the data the practicality and effectiveness of performance assessment developed. The mean value of practicality obtained was 3.48 with the criteria of practicality "High" and does not need to test again. Easy to administer is one of criteria performance assessment have practicality ^[14]. Easy to administer aspect has been conculde in teacher response questionnaire. Performance assessment can improve metacognitive skills by 6.1% and increase the cognitive learning of up to 91.67%. Metacognitive skills increase in the last meeting. Students routinely need to be asked to think about their emerging knowledge and skills, write about what they know and want to know, appreciate how they think, and consider how they can monitor and manage their thoughts and actions-all is aspects of metacognition ^[15]. These statement indicate that metacognitive should be applied routinely in the classroom.

V. CONCLUSION

Performance assessment developed has been validated by an expert assessment and field practitioners with a value of validation is 3.84 with the criteria of validity "Valid". The mean value of practicality obtained was 3.48 with the criteria of practicality "High" and does not need to test again. Performance assessment can improve metacognitive skills by 6.1% and increase the cognitive learning of up to 91.67%.

ACKNOWLEDGEMENT

Thanks a lot to Prof.Dr. Herawati Susilo, M.Sc, P.hD as assessment expert who was validation this performance assessment. Thanks to Dr. Susriyati Mahanal, M.Pd as matter expert who was validation of student's worksheet, and don't forget thanks to Suprpti S.Pd as field practitioners (biology teacher) also student's in X MIA-7 High School 1 Lawang.

REFERENCES

- [1] Saavedra, A.R dan Opfer, V.D. 2012. Teaching and Learning 21st Century Skills: Lessons from the Learning Sciences. Rand Corporation. (Online)
- [2] Handel, M., Artelt, C., Weinert, S. 2013. Assessing Metacognitive Knowledge: Development and Evaluation of a Test Instrument. Journal for Educational Research. (Online), (www.j-e-r-o.com/index.php/jero/article/viewFile/365/172)
- [3] Arifin, Z. 2013. Evaluasi Pembelajaran. Bandung: PT. Remaja Rosdakarya
- [4] Sudjana, N. 2014. Penilaian Hasil Proses Belajar Mengajar. Bandung: PT. Remaja Rosdakarya
- [5] Sunarmi dan Prasetyo, T.I. 2003. Evaluasi Proses dan Hasil Belajar. Malang: Dikti
- [6] Omid, M. dan Sridhar, Y.N. 2012. Effectiveness of Performance Assessment on Metacognitive Skills. Journal of Education and Practice, 3 (10)
- [7] Suwono, H. 2012. Penilaian Hasil Belajar IPA. Malang: Bayumedia Publishing
- [8] Popham, W.J. 2006. Assessment for Education. USA: Pearson Education, Inc
- [9] Hamid, M.S. 2011. Standar Mutu Penilaian dalam Kelas. Yogyakarta: Diva Press
- [10] Sukardi, H.M. 2010. Evaluasi Pendidikan: Prinsip dan Operasionalnya. Jakarta: Bumi Aksara
- [11] Llewellyn, D. 2013. Teaching High School Science Through Inquiry and Argumentation. USA: Corwin A SAGE Company
- [12] Doran, R., Chan, F., Tamir, P. 1998. Assessment. Virginia: NSTA
- [13] Brown, J.H. and Shavelson, R.J. 1996. Assessing Hands-On Science. USA:Corwin Press, Inc
- [14] Greenstein, L. 2012. *Assessing 21st Century Skills*. USA: Corwin Press, Inc

DESIGN STUDENT DEVELOPMENT WORK SHEET (LEARNING CYCLE) 5E TO IMPROVE STUDENT LEARNING OUTCOMES High School CLASS X

Feryco Candra¹, Chandra Ertikanto²

¹Teacher Senior Religious High School ,

² Master of Physics Education, Lampung University

¹feryco_c@yahoo.com

Abstract— This study aims to generate Student Activity Sheet (LKS) -based 5E learning cycle to improve student learning outcomes subjects in this study were students of class X IPA MA Guppi Banjir the school year 2015/2016. The method used by the R & D stage of development following the model of the 4-D by [1] This model consists of four stages of development, define, Design, Develop, and Disseminate. In this article, the study is limited to the design stage (design). Learning learning cycle consists of five stages consisting of generation stage of interest (engagement), exploration, explanation, elaboration, and evaluation. The results of the study concluded that the BLM to improve student learning outcomes learning model learning cycle needed by teachers of physics it can be seen In Figure 1 high school teachers in Waykanan many say yes as much as 64% and as much as 34% and in figure 2 many students say yes as much as 59% and not 41% statement.

Keywords: LKS, learning cycle 5e, learning outcomes.

I. INTRODUCTION

Today the school learning should be centered to student Hence, teachers can develop the learning process, especially learning resource that is able to expose the students' ideas into something valuable and beneficial for him. Source learning has a very important role in the learning process effective and efficient. This was confirmed by the Association for Educational Communications and Technology [2] a source of learning is everything or power that can be utilized by teachers, either separately or in a combined form, for the sake of learning and teaching with the aim of increasing the effectiveness and efficiency of learning objectives, Learningresources linked to the preparation of instructional media.

Learning resources can be obtained from various kinds of instructional media needs. Media is a communication tool that is channeling messages and can stimulate the mind, feelings, and abilities of students so as to encourage the teaching and learning process effective and efficient [3] [4] So the learning media is a tool supporting the learning. With the learning media is expected students will better understand the subject matter they are studying. One type of media that is often used by each school is Student Activity Sheet (LKS). At this time, in the reality of education in the field, many teachers are still widely used every school in the form of worksheets Conventional or worksheets that monotonous, namely worksheets that stay put, stay bought, instant, and without any effort to plan, prepare, and prepare their own [5]. Though teachers know and are aware that the worksheets that they use often does not correspond to the basic competence and indicator. Learning by to use conventional LKS has limitations in enhancing the competence and student characteristics. Materials, questions guidance and tasks in conventional LKS does not correspond to the needs of students and not contextual [6], so it is less improve the competence of students should be increased as optimally as possible. Conventional LKS students do not find referrals that are structured to grasp the material provided.

Based on observations in school worksheets are provided from the school is not the result of the development of the school teachers. However LKS obtained from the publisher that has been provided. With LKS existing learning model was conducted by a monotonous so that the teacher becomes more active (teacher centered). In addition, in a long time, the explanation LKS

with traditional learning models such as the "definition-formula-the-practice-practice examples" it was very easy for teachers but for students it is boring and difficult, thus affecting the learning outcomes of students [7]. This has resulted in students not be able to acquire new knowledge by itself and the learning process is not effective and efficient. It is thus necessary a good learning environment to raise their experience, so that students can gain new knowledge by itself.

In the application of the use of conventional LKS school, learning model used in the learning process is not integrated with LKS used. This has resulted in monotonous learning and students will get bored following the learning process. Therefore, to overcome the disadvantages of conventional LKS LKS needed development in learning physics. At the stage of development of LKS, required conformity existing problems with learning model combined. After studying the conditions of the place and situation of the research, the learning model Learning Cycle (Learning Cycle) is thought to be the right model in the development of LKS. "Learning Cycle are models of how people encounter and acquire new knowledge [8]. Learning Cycle Model is a model of how people discover and acquire new knowledge. The model will invite students to become competent in various aspects, cognitive, affective and psychomotor learning activities.

Learning Cycle-based worksheets, students can find referrals that are structured to grasp the material provided. So that the learning process is student centered. So that students are more active then automatically learning outcomes of students will increase. The purpose of this research is (1) Model-based learning learning cycle in the learning process has not been fully used. (2) LKS to improve learning outcomes of students with learning physics-based learning cycle has not been used. (3) LKS with learning physics model learning cycle takes teachers and students. To overcome this problem, this research aims to design learning student work sheet physics-based learning cycle Cycle learning 5e to improve student learning outcomes.

II. RESEARCH METHODS

Subjects in this study were students of class X IPA MA Guppi Banjit the school year 2015/2016. This study aims to produce teaching materials in the form of worksheets based Learning Cycle 5E to use design and approach to research and development (research & development / R & D). This study was designed to use the design development of the 4D model suggested by [9]. The rationale for this model is due to stage a systematic and more detailed. Learning learning cycle consists of five stages [10]. which consists of the generation stage of interest (engagement), exploration, explanation, elaboration, and evaluation.

III. RESULTS AND DISCUSSION

On this occasion will be presented some of the things found in the field based on the observations made during the field study.

A. Teacher needs analysis questionnaire

questionnaires given to the needs of teachers in high school -teacher teachers in waykanan much as 5 teacher where the questionnaire contains 10 statements about the information related to the learning process at school. Here are the results of the answers to a questionnaire given to teachers.

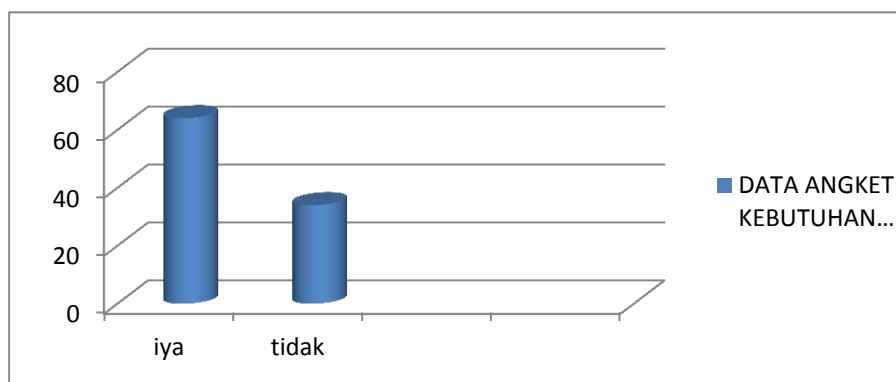


Figure 1. Results of giving questionnaires to teachers

In Figure 1 high school teachers in Waykanan many say yes as much as 64% and as much as 34% it means LKS required by existing teachers -Teacher diwaykanan. Then the understanding of the function and role in learning LKS, LKS only limited use as a material for the provision of duty and the content contained in the LKS is only a summary of the material and practice questions. LKS should be used to activate students in learning activities, helping students find and develop concepts, train students to find a concept, an alternative way of presenting the subject matter which emphasizes the involvement of the student, and can motivate students [11].

In addition LKS should also be able to play a part in practical activities in the laboratory. With the help of student worksheets can be more convenient and effective in gathering information / data so that students can draw conclusions or provide answers to questions problem (formulation of the problem) rationally because of an underlying information / data that can be accounted for, not only by the argument [12]

B. 2. Student Needs Analysis Questionnaire

Questionnaires given to the needs of students in the class X as many as 30 students in which the questionnaire contains 10 statements about the information related to the learning process at school. Here are the results of the answers to a questionnaire given to students.

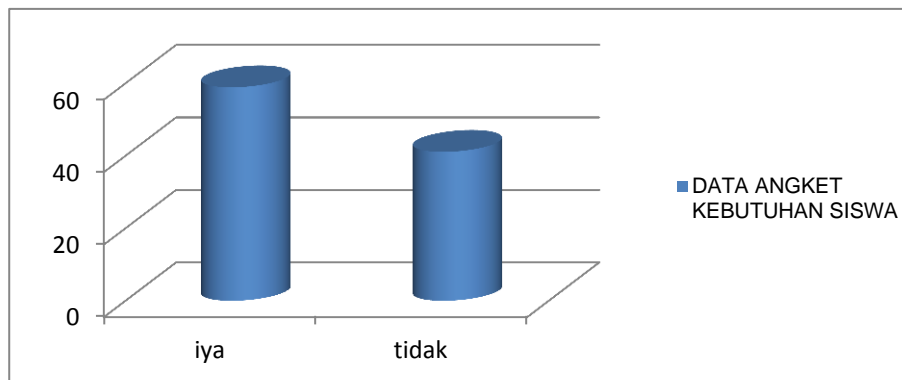


Figure 2. FIGURE 2. The results of the questionnaire to the student administration

In figure 2, it can be seen that in the learning of physics required learning by to use worksheets that can be seen from the image many students say yes as much as 59% and the statement is not 41% to use the worksheets to the model learning cycle, in Figure 2 also found that students own LKS as used by teachers but some students felt that the presence of LKS not feel the ease in understanding the subject matter physics .This disebabkan for LKS used only serves as a means of exercise do the questions and complete the tasks assigned by the teacher. Should benefit their worksheets for the students is that learning to use open LKS can improve student learning outcomes.

IV. CONCLUSION

Based on the results of research and discussion can be summarized as follows: (1) Model-based learning learning cycle in the learning process has not been fully used. (2) LKS to improve learning outcomes of students with learning physics-based lerning cycle has not been used. (3) LKS with learning physics model lerning cycle takes teachers and students.

ACKNOWLEDGMENT

This project would not be possible without the help from members of our team who are not author on this paper Erlida Amnie, Yuda Seta Mahendra, Saiful Ali Nurdin, F Bayu Nirwana, Iwan Setiawan and Novinta Nurul Sari.

REFERENCES

- [1] S,Thiagarajan & D.S ,Semmel. (1974) Instructional Development for Training Teachers of Exceptional Children: A Sourcebook. Minneapolis: University of Minnesota
- [2] Depdiknas 2008.Panduan Pengembangan Bahan Ajar. Jakarta: Direktorat Pembinaan Sekolah Menengah Atas.
- [3] A, Arsyad. 2012. Media Pembelajaran. Jakarta: PT Raja Grafindo Persada.

- [4] M.B, Usman. dan Asnawir. 2002. Media Pembelajaran. Jakarta : Delia Citra Utama.
- [5] A, Prastowo. 2012. Panduan Kreatif Membuat Bahan Ajar Inovatif .Yogyakarta Diva Press
- [6] A, Prastowo. 2012. Panduan Kreatif Membuat Bahan Ajar Inovatif .Yogyakarta Diva Press
- [7] K,Yenilmez. and M, Ersoy. (2008). "Opinions Of Mathematics Teacher Candidates Towards Applying 7e Instructional Model On Computer Aided Instruction Environments". Internasional Journal of Instruction Volume 1 No. 1. Hal 49-60.
- [8] Abruscasto, Joseph dan A, Derosa Donald. 2010. Teaching Children Science A Discovery Approach,7 th. Boston: Allyn dan Bacon.
- [9] S, Thiagarajan. & D.S. Semmel. (1974) Instructional Development for Training Teachers of Exceptional Children: A Sourcebook. Minneapolis: University of Minnesota.
- [10] Lorschbach. 2002. The Learning Cycle as a Tool for Planning Science Instruction.
- [11] Trianto. 2010. Mendesain Model Pembelajaran Inovatif-Progresif (Konsep, Landasan, dan implementasinya pada Kurikulum Tingkat Satuan Pendidikan (KTSP). Jakarta: Kencana.
- [12] M, Hosnan. 2014. Pendekatan Saintifik dan Konstektual dalam Pembelajaran Abad 21. Bogor: Galia Indonesia.

Development of Vocational Interest Scale: A preliminary study of the psychometrics properties*

Firmanto Adi Nurcahyo

Psychology Faculty, Pelita Harapan Surabaya University
firmanto.adicahyo@uph.edu

Abstract—Vocational interests lead to certain types of work that suits for some people. The disclosure of interest is an important matter considering the contributions of interest to education and employment area is appreciable. Estimating the individual interests can be done based on individual scores of vocational interests scale. This study is a preliminary study of a research that aims to develop vocational interest scale for adolescent in Indonesia. In particular, this study aims to conduct item analysis using item response theory based on preliminary data. The scale was developed based on Holland's interest theory that distinguishes 6 types of interest, which are Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Each interest is characterized by the selection of certain activities and personal traits. The scale consists of 24 items with three response options representing three types of interest. Subjects were asked to select two options that suit themselves. Subjects were 120 students of the Faculty of Psychology. The strength of the link between interests theoretically is the basis for the scoring. The strongest linkage between two types of interest is manifested in the option of a and b. Subjects who choose those options will receive a score of 3. Item analysis was performed using WINSTEPS. Results showed that some items still need some improvement. This is shown by a number of items that contain DIF. Limited number of respondents become the weakness of this study. This suggests that the study still needs further improvement.

Keywords: vocational interest, scale, item analysis

**This study is part of doctoral research internship supervised by Prof. Dr. Badrun Kartowagiran.*

I. INTRODUCTION

Vocational interests lead to certain types of work that suits for some people. Sugiyanto said that interest plays an important role in determining the course of study [1]. The disclosure of interest is an important matter considering the contributions of interest to education and employment area is appreciable. However, only a few people were able to recognize their interest. As a result, many individuals experience confusion, for example in determining the field of education. Sawitri noted the same problems occur from year to year where high school graduates do not know where they have to continue their studies [2].

Biological condition of the individual, as well as early experiences led to the selection of a particular activity. Selection of the specific activity is then developed into an interest. Fulfillment of interest makes individual feel satisfied. Efforts of the fulfillment of interest will lead to the development of certain abilities and will be followed by crystallization of the values associated with the interests [3].

Holland has distinguished individual interests into 6 types namely Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC) [3]. Individuals with R type prefer activities that involve a clear and systematic treatment of the particular object, such as equipment, machinery, and animals, as well as avoiding activities related to education or medication. Individuals with I type prefer activities that are observational, symbolic, and systematic, as well as research on phenomenon, as an effort to understand and to control the phenomenon. Individuals with A types prefer activities that are free, unsystematic, and ambiguous and avoid activities that are clear, systematic and well structured. Individuals with S types prefer activities that involve the treatment of others in terms of providing information, training, developing, curing, and explaining something. These individuals do not like structured activities, that involving materials, equipment, and machinery. Individuals with E types prefer activities that involve the treatment of others to achieve organizational goals or economic gain.

These individuals do not like systematic, symbolic, and observational activities. Individuals with C type prefer activities that involve the treatment of the data in a systematic, explicit and structured. These individuals do not like ambiguous, as well as unsystematic activities.

The relationship between each type of interest can be illustrated through a hexagonal pattern (Figure 1) [3]. The shorter the distance between two types of interest, the greater the similarity possessed. R and I have a short distance, indicating that those types have high similarity. This is in contrast to I and E that have longer distance. Both types of interest are very different from one another. I and S have a moderate distance. Therefore, there is a moderate similarity between those types of interests.

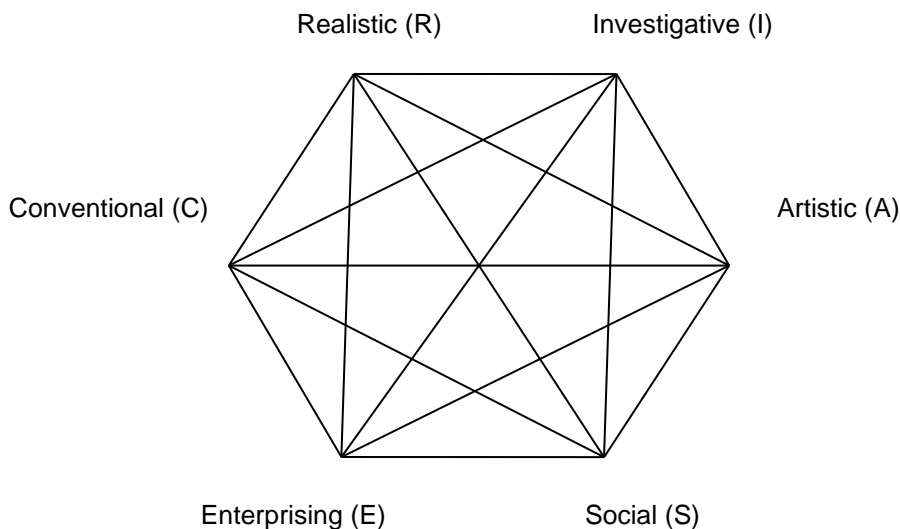


Figure 1. Holland's Hexagonal Model [3]

It takes a measure to gain an understanding of interest. Measurement of interest can be performed with a vocational interest scale. Holland said that the estimation of individual interests could be done based on individual score of a vocational interest scale [3]. The results of these scores provide an overview of individual vocational interests.

Self Directed Search (SDS) is an interest scale, developed by Holland that is widely used today. SDS was published in 1970 and revised in 1977, 1985 and 1994 [4]. Psychological Assessment Resources revealed that over 30 million people have used SDS [4].

Although it has been widely used, the SDS is not free from drawbacks. One of the problems found is the content validity issue. Ting found some problems in job content items [5]. This was demonstrated by the difficulties of the students in understanding some of the jobs in the SDS. Internal structure validation is another problem. Although most studies support Holland's theory, there is still a debate about the application [6]. The problem that arises is the sample of Americans demonstrated support for the structure based on Holland's theory compared to samples from outside the United States.

To overcome these problems, development of interest scale is needed. Furr explained that scale development includes four stages as shown in figure 2 [7]. The first stage can be realized by defining the variable to be measured as well as differentiating this variable from other similar construct. The first stage followed by the second stage, which is writing a number of items that are relevant to the intended construct. Item writing process is based on the clarity of the construct and its relevance. The third stage is the stage where the items are given to respondents who represent the target population. This stage produces data that will be used to evaluate the psychometric properties and the quality of the items. The fourth stage is the stage where the reliability and validity of the scale was obtained. The result of these stages is a scale with good psychometric quality.

This study is a preliminary study of a research that aims to develop vocational interest scale for adolescent in Indonesia. In particular, this study aims to conduct item analysis using item response theory based on preliminary data. The results of the item analysis will be taken into

consideration, whether such items can be used in the scale of interest, or still needs further improvement. Selected items are items that have good psychometric qualities.

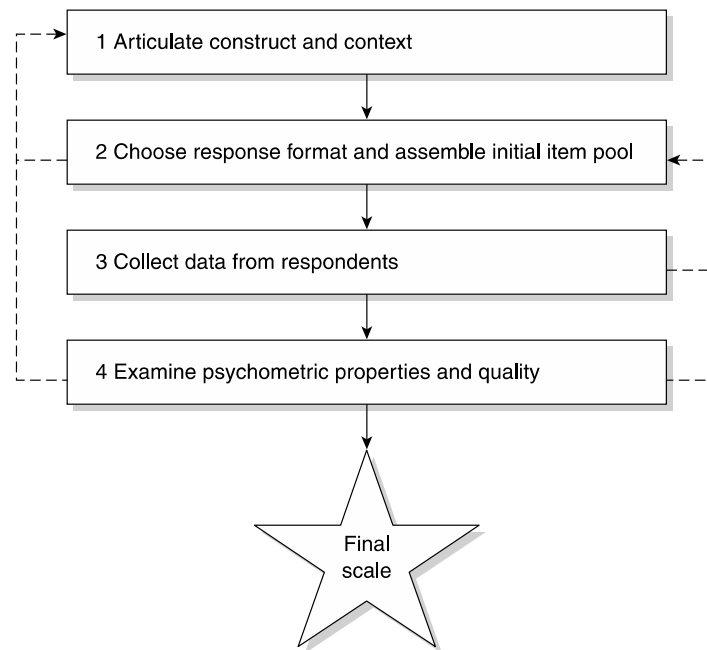


Figure 2. Scale development [7]

II. METHODS

A. Respondents

Respondents of this study were 120 students of the Faculty of Psychology. 38% of respondents are male and 62% are female. Most of the respondents were first year students.

B. Instrument

The scale consists of 24 items with three response options representing three types of interest. For each item, respondents were asked to choose two options. The strength of the link between interests theoretically is the basis for the scoring (see Figure 1). The strongest linkage between two types of interest is manifested in a and b option for each item. Subjects who choose those options will receive a score of 3. The example of the item can be seen below (Item 2).

In doing my work, I feel I have the ability to:

- a. lead and coordinate others
- b. communicate with new member
- c. operate tool to support thw work

Item 2 asks about respondent's ability to work, with three options of activities. Option a represents E type, option b represents S type, and option c represents R type. From figure 1 it can be seen that E and S have short distance, E and R moderate distance, and S and R long distance. Therefore the scoring will be 3, if respondents choose a and b options, 2 if respondents choose a and c options, and 1 if respondents choose b and c options.

The scale was developed based on the psychology profession in accordance with the course of study taken by the respondents. According to Holland, profession can be characterized by the profile (with a three-letter code) that indicates the dominance of the six types of interests. Social-Investigative-Artistic (SIA) is considered as the dominant code for the profession of psychology [3]. Respondents will have a high score, if he choose options which represents those types of interest.

C. Data Analysis

Factor analysis was conducted to determine the dimensionality of the scale. Therefore, confirmatory factor analysis was performed using AMOS. The criteria used to assess the fit between the theoretical model to the data is based on Chi-square value with $p \geq 0.05$, Goodness of Fit Index (GFI) ≥ 0.9 , Root Mean Square Error of Approximation (RMSEA) ≤ 0.05 [8].

Item analysis was conducted by using Item Response Theory (IRT) that is applied to polytomous data. WINSTEPS program, which based on IRT one parameter (Rasch Model), was used to perform data analysis. Items measure output shows the logistics value of each item. Item with high logistic value indicates item that is difficult to be approved, and vice versa.

The analysis of differential item functioning (DIF) was conducted to test item bias. An item shows DIF if individuals with similar abilities who have different opportunities to answer the item correctly [9]. In this study DIF refers to the difference in the probability of the subject in approving an item, even though they have a similar level of interest. The difference in this case is in terms of gender. WINSTEPS program was also used to detect items that are affected DIF by gender.

III. RESULT AND DISCUSSION

CFA was conducted to determine whether the scale is unidimensional, as required by Item Response Theory. AMOS program was used to conduct CFA. The results showed that there are five items which have negative regression weights. Those five items were item 5, 6, 8, 18, and 23. This is certainly to be something that was not expected, considering all of the items should have positive weights on the latent trait. Therefore, those five items were then removed from the scale and CFA was conducted again with the remaining items. Result of one factor model showed CMIN=178.902, $p=0.067$. This confirms that the model is unidimensional. This is also supported by the value of RMSEA=0.039, while GFI=0.873.

Unidimensionality was also confirmed by using WINSTEPS output, which is table of standardized residual variance. It can be seen that the percentage of raw variance explained by measured is 21.1%. The percentage which is above 20% indicates that minimum unidimensionality requirements can be met [10].

Result of WINSTEPS in the summary table shows that the average of person measure is 0.33. The average value of more than 0.0 indicates a tendency that most respondents answered agree. This is certainly a good thing, considering the respondents who are students of psychology, tend to approve items that led to an interest in psychology. However, the result of person reliability shows 0.5. This means that the consistency of the answers of the respondents is still weak. This is quite different from the results of item reliability which is 0.95. This indicates that the quality of the items on the interest scale is good.

Based on Item Measure output, it can be seen that item 7, with logistic value of 1.35, is the most difficult item to be approved. On the other hand, the easiest item to be approved is item 3 (logistic value =-1.77). Misfit items are detected based on the value INFIT MNSQ, obtained from the item fit order table. Items that have INFIT MNSQ value greater than the sum of the mean and standard deviation, indicates the item is misfit. Based on this it can be seen that the misfit items are item 7, 13, and 20, which have INFIT MNSQ value 1.51, 1.30, and 1.24. All of which are greater than the sum of the mean and standard deviation, which is 1.23.

Ratings scale analysis was conducted to test whether the option used rank in the scale of interest is confusing the respondent. The interest scale was developed using three options, which were scored from 1 to 3. In the rating scale table of WINSTEPS output, it can be seen that there are three categories of scoring which refers to 1 to 3. Those three categories were then followed by the observed averages which are -0.13, 0.26, and 0.66. The increased observed average value indicates that respondents can understand the level of the given options.

The detection of item bias is conducted based on the output of items DIF table. In this study, item bias is detected based on gender. The item probability that is less than 5%, indicates that the item is biased [10]. There are three items that have a probability value of less than 5% i.e. items 1, 10, and 15. Figure 3 shows items plot based on gender.

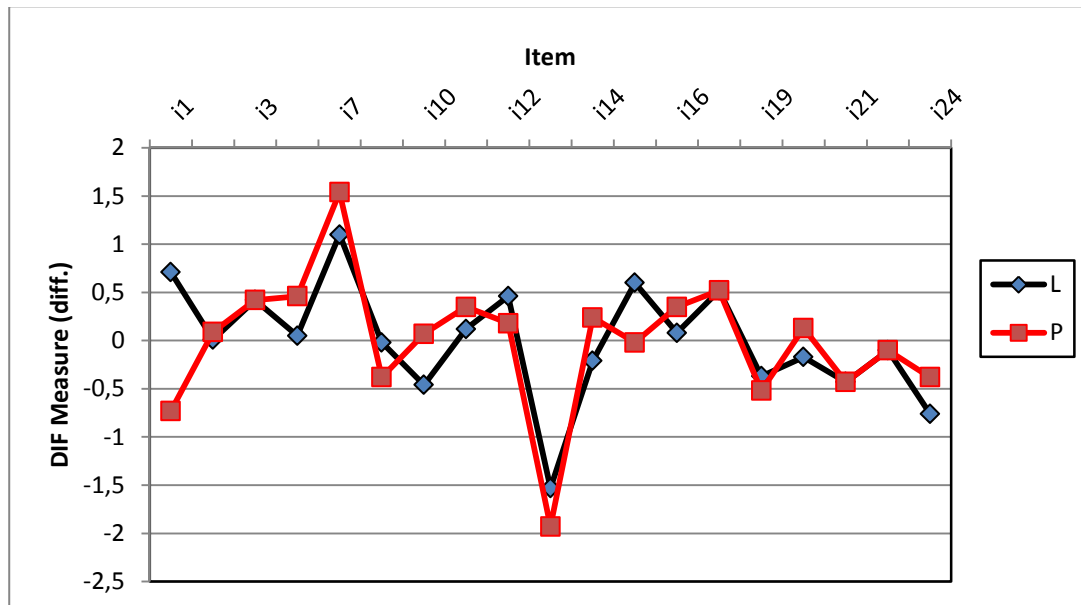


Figure 3. Items Plot Based on Gender (L=Male, P=Female)

It can be seen from the plot that item 1 and 15 are likely to favor male than female. This is in contrast with item 10 that actually looks favorable among female than male. Therefore, the three items need to be checked along with the accompanying options.

Item 1

The topic of reading that interests me:

- Literary or art
- Healing of an illness
- Sports or related to automotive

The c option seems to be easier chosen by male than female. Although that option reflects realistic interest, such option is generally easier approved by male than female.

Item 10

In an event committee, I feel I can work best when I work as:

- contact person for anyone who wants to ask the event
- take care of the decoration and documentation
- take care of everything related to record notes and finance

Options a and c seems easier to be approved by female than male, since both these options normally performed by female.

ITEM 15

In this life, I am interested to have:

- high creativity
- extensive social relations
- good physical condition

Item 15 has the same problem with item 1. The c option seems to be an easier chosen by male than female, although it reflects realistic interest.

The study shows that despite in general the interest scale has good reliability, some items are still require further improvement. Improvements made to the items that experienced misfit, as well as items that are detected DIF.

Limited number of respondents becomes the weakness of this study. Another limitation, the interest scale developed in this study is limited to the work of psychology, which has specificity in Social Investigative and Artistic interest. Therefore, the development scale of interest that can be used to estimate interest in other occupations, are still needed.

REFERENCES

- [1] Sugiyanto, "Alasan Pemilihan Jurusan Studi Siswa SMA Kelas 1." Universitas Gadjah Mada Yogyakarta, 1983.
- [2] D. R. Sawitri, "Pengaruh Status Identitas dan Efikasi Diri Keputusan Karir terhadap Keraguan Mengambil Keputusan Karir pada mahasiswa Tahun Pertama di Universitas Diponegoro," *J. Psikol. Undip*, vol. 5, no. 2, 2009.
- [3] J. L. Holland, *Making Vocational Choices: a theory of vocational personalities and work environments*. USA: Prentice Hall, 1985.
- [4] V. C. Dozier, J. P. Sampson, J. G. Lenz, G. W. Peterson, and R. C. Reardon, "The Impact of the Self-Directed Search Form R Internet Version on Counselor-Free Career Exploration," *J. Career Assess.*, vol. 23, no. 2, pp. 210–224, May 2015.
- [5] S.-M. R. Ting, "Comparing the Chinese Career Key and the Self-Directed Search With High School Girls in Hong Kong," *J. Career Dev.*, vol. 34, no. 2, pp. 192–209, Dec. 2007.
- [6] Y.-T. Sung, Y.-W. Cheng, and J.-S. Wu, "Constructing a Situation-Based Career Interest Assessment for Junior High School Students and Examining Their Interest Structure," *J. Career Assess.*, Apr. 2015.
- [7] R. M. Furr, *Scale construction and psychometrics for social and personality psychology*. Los Angeles ; London: SAGE, 2011.
- [8] J. Arbuckle, *Amos 16.0 user's guide*. Chicago, Ill.: SPSS Inc., 2007.
- [9] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, *Fundamentals of item response theory*. Newbury Park, Calif: Sage Publications, 1991.
- [10] B. Sumintono and W. Widhiarso, *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: Trim Komunikata Publishing House, 2013.

Contextual Approach Using Pictures as a Media Increased Result and Motivation of Mathematical Learning

(Mathematical Learning of Fractional Addition by Equalizing the Denominator)

¹Ihsana El Khuluqo, ²Ningrum Rosyidah

¹Education Administration Management, University of Prof. Dr. Hamka

²Teacher Of Mathematic, SDIT AL KHALIFA

ihsana_khuluqo@uhamka.ac.id

Abstract—This research aims to understand the result of mathematical learning of Class IV student in elementary school. (1) Discovering the influence of the Contextual Approach to the student's motivation. (2) Discovering whether the Contextual Approach can increase student's motivation by equalizing the denominator for the class IV students. The Research uses descriptive method by using inferential statistical analysis technique. "Descriptive method is a method that focuses to solving current and actual issues. Inferential statistical technique is a statistical method to analyses sample data and creating general inference for the population." (Nisfiannoor, 2009:4). The approach used by the research is the quantitative approach. Quantitative approach is a research approach that uses statistical technique to test hypotheses and research instrument to be used to uncover data in a certain measurement scale, to be able to create a generalized statement for the whole population. (Sudjana and Ibrahim, 2004:8). There are several conclusions to be inferred from the result of the research. From the cyclical analysis, the improvement of the learning capacity was successfully implemented in the cycle I, cycle II, or even cycle III. Cycle I experienced an increase in average class scores from 42 to 72. Cycle II also experienced increasing from 72 to 81. While Cycle III experienced increasing from 81 to 83. The increases in scores are due to the improvement in the teacher's teaching method by implementing contextual approach. The approach ultimately increases the students' enjoyment of the learning, leading to an increase in the motivation for learning, therefore increasing the result of the learning itself. If the learning result improves, the quality of the education in the school also improves.

Keywords: Learning, Education, Contextual Approach

I. INTRODUCTION

Education is one of the most fundamental needs of mankind, for it was education that will prepare a man to face problems and challenges in his life. Therefore the quality of education is linearly related with humans' quality of life, an increase in the quality of education is an increase in the quality of life. In order to increase the quality of education there is a need for every stakeholder to participate in the quality improvement endeavor, most of all teachers as the educator. Not only as an educator, a teacher has multiple roles in the education process, a teacher must be able to pass on knowledge, facilitate learning, guide, and motivate the student to actualize the potential inherent in each and every student. As a motivator, teacher has to be able to drive the student to learn, to stimulate the students' curiosity and willingness to learn. This role is one of the most important roles a teacher has in order to improve students' learning motivation and capacity.

In order to motivate the student, a teacher has to have a mastery over the methods and models of teaching. The sheer number of students with different condition and capabilities means that there is no one model that is suitable to teach and motivate the entire student, therefore the ability to choose and implement a suitable method is a must for a teacher. However, in practice most teacher still dominate the education process, making the whole process teacher centric as opposed to the student centric ideal that is more suitable to increase the potential of learning in the students. The results are the student gets tired and bored of the whole education process.

The incessant memorizing, and the monotony of the education process severely inhibit the learning capacity of the student, thus lowering their test result.

Regarding said problem, this research tries to find out the appropriate method to increase the learning motivation and result among class IV student in elementary school. According to a latest research, the result of the mathematical learning particularly on the Fractional Addition in class IV student is still below the standard imposed by the government, achieving only 50 on average while the standard is 60. Therefore there is a need to improve the educational quality by special means.

A. Research Question

How to increase the result of the mathematical learning process, particularly the Fractional Addition, by equalizing the denominator using picture as a media in the contextual approach in Kapuk Muara Pagi 07 Elementary School, 2014/2015 academic year Kapuk Muara Jakarta Utara?

B. Research Aim

- To discover the influence of contextual approach to students' learning capacity
- To discover whether the usage of contextual approach can increase students' motivation

II. RESEARCH METHOD

This Research uses descriptive method. "Descriptive method is a method that focuses to solving current and actual issues." (Nisfiannoor, 2009:4)

A. Data Collecting Technique

Data collecting technique used in this research are:

- Learning product test that are used to gather data regarding the result of Class IV mathematical learning process. The tests are an essay question with the answer key and evaluation criteria.
- Observation is used to gather data on the educational activity.

Gathered data will be analyzed descriptively and using percentage technique to uncover the learning pattern in the mathematical learning process. The calculation is as follows:

a. Mathematical Learning Product

Calculated by analyzing the average of the formative test done:

$$\text{Score} = \frac{\text{Total Students Score}}{\text{Number of Students}} \times 100\%$$

b. Students Aptitude

Calculated by averaging the formative test score with the number of the question in the test:

$$\text{Score} = \frac{\text{Total Students Score}}{\text{Number of questions}} \times 100\%$$

B. Analysis technique

The technique used by the researcher is "Inferential statistical technique is a statistical method to analyses sample data and creating general inference for the population." (Nisfiannoor, 2009:4). The approach used by the research is the quantitative approach. Quantitative approach is a research approach that uses statistical technique to test hypotheses and research instrument to be used to uncover data in a certain measurement scale, to be able to create a generalized statement for the whole population. (Sudjana and Ibrahim, 2004:8).

III. LITERATURE STUDY

A. Learning Result

1) Learning

According to Winkel in Purwanto, learning is a psychological activity that occurs in an active interaction with the environment that results in changes in knowledge, skill, and attitude. Sardiman A.M. further expand on the definition by providing the essential principles of learning:

(1) Learning is connected to the human potential and attitude; (2) Learning need stages and a certain maturity of the student; (3) Learning is more effective with intrinsic motivation rather than learning because of fear and stress; (4) Learning mostly consist of trial and error, and conditioning; (5) students' learning capacity must be calculated to determine the appropriate learning content; (6) there are 3 ways of learning: Tutoring, Contact and experience, Imitating; (7) Learning through practice and direct experience is much more effective than memorizing; (8) students' experience influence students' learning capacity (9) Meaningful lesson is easier to absorb than meaningless lesson; (10) Information of students conduct, behavior, achievement, and failings can stimulate students motivation; (11) Learning process must be as varied as possible in order to stimulate students' experience and motivation.

2) Learning Product

Learning product is the result of the whole educational process. In order to achieve good result the education process need to be able to sustain a fun and inspired environment for the students. According to Purwanto learning process can be explained etymologically by understanding the word "Learning" and "Product". Product refers to the output of a given activity or process that cause a functional change in the input. Learning is an activity aimed to create changes in the learning individuals' attitude. The change in the individuals' attitude is the learning product. Winkel in Purwanto states that learning product is a change that caused human to alter its behavior. In that context-learning product is the end result of the learning process in line with the ends being attained. Sri Anitah suggested that there are 2 factors that determine the learning product; internal factor (motivation, goals, talent, etc.) and external factor (environment, family, school program, teacher, etc.)

B. Mathematics

Mathematics and Fractions according to experts

According to Johnson and Rising, Mathematics is a mode of thinking, organizing, and logical proofing using a symbolic language that is defined clearly, accurately, and succinctly. Mathematic is a symbolic language that expresses quantitative and spatial relationship that simplifies and makes thinking congenial and convenient as Johnson and Myklebust believes.

Fraction is a number that consist of numerator and denominator that represent a part of the whole. Fraction described how many parts of a certain whole are there in a conversation or in a mathematical exercise. Usually written down as e.g. $\frac{3}{4}$, (three – quarter), $\frac{1}{8}$ (one – eight).

C. Contextual Teaching and Learning

Contextual Teaching and Learning (CTL) is a holistic education process that aims to make lessons congenial and convenient for student to understand by relating the lessons with students' experiences. It engages students in meaningful, interactive, and collaborative activities that support them in becoming self-regulated learners. In relating the lessons with real work experience of the students' there are 7 main component of effective learning that must be kept in mind. Those components are: (1) Constructivism; (2) Questioning; (3) Inquiry; (4) Learning Community; (5) Modeling; (6) Reflection; and (7) Authentic Assessment.

3) Key Concepts

Contextual is one of the key concepts that enable student to learn meaningfully. By learning meaningfully, students are more apt to apply the lessons in their everyday life, therefore achieving the goal of education. There are 9 key context that concern a student: (1) Goal Context, what goal is desired?; (2) Lessons Context, what lessons is given?; (3) Source Context what source is appropriate?; (4) Target Context, who is the target recipient?; (5) Teacher Context, who is doing the teaching?; (6) Method Context, what method is used?; (7) Product Context, how will the product measured?; (8) Maturity Context, is the recipient ready for the lessons? (9) Environment Context, what kind environment is conducive to learning?

As mentioned above, CTL is concerned in engaging students in meaningful, interactive, and collaborative activities that support them in becoming self-regulated learners by relating the lessons into real world context. As such there are 5 characteristics of the approach: (1) activating prior knowledge; (2) acquiring new knowledge; (3) understanding said knowledge; (4) applying the knowledge in the appropriate context; (5) reflecting upon the knowledge.

4) CTL Indicators

As it mentioned above there are 7 key components in the CTL. Those are:

- Constructivism: constructing knowledge from students' real world personal experience and personal knowledge structure
- Inquiry: Main activity of CTL, the students undertakes the process of Observation, Questioning, Hypothesize, Data Gathering, and Conclusion to uncover new knowledge.
- Questioning: A useful strategy of contextual learning to discover new knowledge and stimulating the students' motivation and understanding
- Learning Community: a community of shared learning, created by grouping together students to encourage communications between them
- Modeling: modeling is enunciating or expressing abstract thought into words. In CTL modeling doesn't have to come from teacher, the students or third party can be actively participating.
- Reflection: reflecting is revisiting prior experience and contemplating them. Its implementation in the CTL is that the teacher provides time for question by students on what did they learn today
- Authentic Assessment: it is a process to collect data relevant to the students learning condition. In CTL the learning condition must be understood by teacher to make sure the students are learning. The focus on the assessment can be on the output or the whole process

D. CTL Forms

There are 5 main forms of CTL. The first is **relating**, in which the teacher relate the content of the lessons into some prior knowledge of the students. Second form is **experiencing**, that is learning by personally experiencing the lessons, through exploration, discovery, and invention. Hands-on activities can involve manipulative, problem-solving activities, and/or laboratories or projects Third form is **applying**, in which the student apply what he knows in a problem solving activities. The teacher can encourage the behavior by giving them realistic and relevant problem. Fourth is **cooperating** in which the students cooperate in a group to solve a given problem, it also encourage the cooperative spirits in the students. Fifth is **transferring**, in which the teacher transfer their knowledge by creating learning experience focused on understanding rather than memorizing.

E. Pros and Cons

Every model always has its advantage and disadvantages, as with the CTL. The advantages of CTL are that the lessons are much more relatable; the students understand the lessons rather than memorizing it; it also encourage critical thinking and problem solving attitude coupled with the courage to speak up their mind, therefore helping them to increase their interpersonal skill. On the other hand, the CTL approach necessitates a large investment in time and resource to make the model works.

F. CTL Criteria

As a summary, the criteria of a CTL based method are as such: the education process are student centric in which the students actively understand knowledge by working together in a group exercises while relating the knowledge to real world practical experiences. The whole learning processes are founded on prior experiences, in which learning is based on personal initiative, the student become self-regulated learner that learn purely for his own self-satisfaction and wellbeing rather than through fear. As such the knowledge and skills of the students expands according to their own experiences.

G. Picture Media

Etymologically, media is taken from the Latin word "*medium*" that literally means a connector or transmitter. Therefore a picture media is a visual or graphical medium that transmits a certain idea through usage of imagery and words (Sudjana, 2007: 68). Picture media can also be understood as an expression of event or experiences through graphical expression such as lines, words, symbol, or pictures (Azhar Arsyad. 1995: 83). According to Gagne (Arief S. Sadiman, 2007: 6), picture media is among the most widely used media in learning, because its ubiquity and ease of transmitting the idea.

H. Characteristic of Class IV Students in Elementary School

Essentially, elementary school students are the same as any other students. They are polite, friendly, good hearted, and obedient to the teacher. Students are happy to go to school because beside they get to learn science and general knowledge; they also get to learn religious knowledge. Most of the elementary school graduate goes on to enroll onto higher school. Thanks to teachers motivating students, the students get to enroll into desired school.

IV. RESEARCH RESULT

A. Result of Increase in Educational Quality

From the data and the result of the analysis, it was found that there is improvement between learning product cycles, be it between cycle I to cycle II, or cycle II and cycle III. The evaluation results from the first cycle shows an improvement, from the average score of 42 pre cycle, to average score 72 in the end of the first cycle. The second cycle experiences the same improvement from 72 to 81 averages. While on the third cycle, the improvement peaks at 81 to 83. The improvements were caused by the usage of CTL with picture media as an improvement to the teaching methods.

Fun and inspired environment of learning can increase students' motivation, in effect increasing the students' learning product. If the increase of the learning product held out, in the long run, the overall quality of the educational process also experience an increase.

The processed data on the increase of learning product in elementary school student class IV is shown in the following graph:

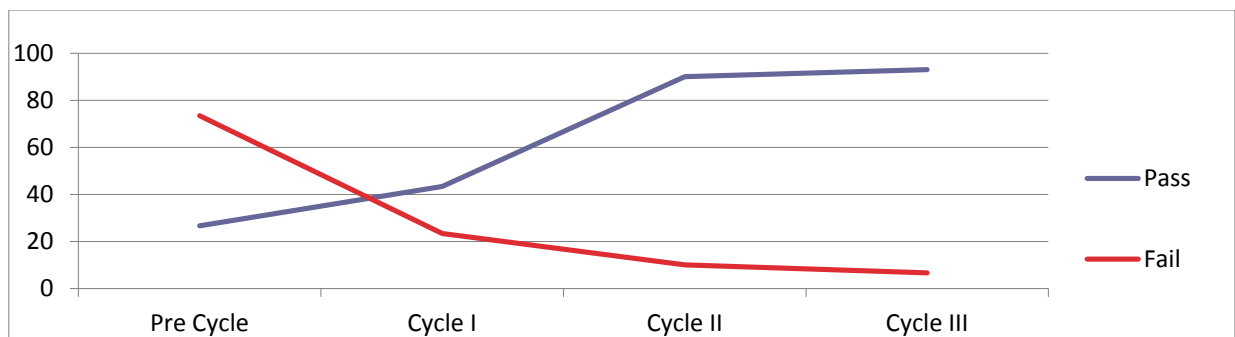


Figure 1. Formative mathematic test result of elementary school class IV Pre Cycle, Cycle I, Cycle II, and Cycle III

V. CONCLUSION AND REMARK

A. Conclusion

Generally the implementation of the three-improvement cycle is essential in improving the learning process of mathematics, especially on the subject of fractional addition by equalizing the denominator, in the second semester of class IV. Departing from previous problems, which will be solved through reflection and observation by supervisors.

With a good cooperation between teachers, principal, and the tutor, CTL with picture media can and will increase students' enthusiasm, motivation, and learning product of Mathematical Learning of Fractional Addition In Elementary School Class IV by Equalizing the Denominator. As proven by the observation increase in average class scores from 42 to 72 in Cycle I, from 72 to 81 in Cycle II, and 81 to 83 in cycle III.

B. Remark

Ultimately it can be mentioned conclusions; there are several points of importance to be done by teachers to improve the learning process:

- A teacher has to be able to stimulate the students to take active role in the learning process.
- A teacher has to be able to choose the appropriate method, media, and model of teaching that is suitable and fun for the lessons at hand.

- A teacher has to be able to provide a wide variety of test and problem in order to evaluate the students' progress.
- A Teacher has to always strive to improve his teaching ability and skill, in designing, implementing, evaluating, and improving a learning process.

REFERENCES

- [1] Makmun, (1996), *Psikologi Pendidikan*, Bandung: Rosdakarya.
- [2] Sardiman, A.M, (2011), *Interaksi & Motivasi Belajar Mengajar*, Jakarta: Pt. Raja Grafindo Persada.
- [3] Rusman, (2012), *Model – Model Pembelajaran*, Jakarta: Pt. Raja Grafindo Persada.
- [4] Anitah, Sri. (2009). Modul 3.Strategi Pembelajaran: *Jenis-jenis Metode Mengajar*. Jakarta: Universitas Terbuka.
- [5] Anitah, Sri. (2009). Modul 2.Strategi Pembelajaran: *Pengertian Belajar*. Jakarta: Universitas Terbuka.
- [6] Ruhimat, Toto dan Hernawan, Asep.(2009). Modul 5.Strategi Pembelajaran: *Jenis-jenis Metode Mengajar* .Jakarta: Universitas Terbuka.
- [7] Taniredja, Tukiran, dan Efi Miftah, dkk. (2011), *Model – Model Pembelajaran*
- [8] Dedi, (2013). Pengertian Pembelajaran Kontekstual, retrieved on April 28, 2016 from <http://dedi26.blogspot.com/2013/06/pengertian-pembelajaran-kontekstual.html>
- [9] *Inovatif* , Jakarta: ALFABETA.
- [10] <http://penkont.blogspot.com/2013/11/daftar-isi-toc-o-h-z-u-kata-pengantar.html>
- [11] (<http://kiflipaputungan.wordpress.com/2010/06/27/pengertianmedia pembelajaran>)

THE CONTENT VALIDITY OF THE EVALUATION MODEL IN THE AFFECTIVE DOMAIN IN ISLAMIC EDUCATION INSTRUMENTS

Iskandar Tsani
STAIN Kediri

Abstract- The evaluation model in the affective domain in Islamic education consists of three dimensions, which are input, process and output. The article aims to examine each dimension of evaluation focusing on the assessment of its content validity. For this purpose, focus group discussions have been conducted by involving a number of experts in related fields, i.e. Islamic education, measurement, evaluation, and linguistics. The instruments were measured by using assessment forms the gradation of which are "excellent", "good", "adequate", and "less". Deploying the Aiken's formula, the analysis of the content validity showed that the content validity index of input construct is 0.942, $p=0.002$, teacher competency is 0.882, $p=0.006$, means of education is 0.865, $p=0.008$, curriculum is 0.872, $p=0.007$, process construct is 0.944, $p=0.001$, output is 0.945, $p=0.001$, discipline is 0.886 $p=0.005$, respect to others is 0.898, $p=0.004$, responsibility is 0.876, $p=0.006$, obedience is 0.871, $p=0.007$. From the numbers it can be concluded that the content validity of the evaluation model in the affective domain in Islamic education instruments are categorized as good.

Key words: *affective domain, content validity, evaluation model,*

I. INTRODUCTION

The current trends show us that assessment of religiosity among students is more heavily based on to what extent they achieve cognitive understanding of their religions. Religious education is evaluated using the established standards in which the results are given in numeric figures. It is also common that outside the learning process in schools, preservation of religious rituals such as fasting and five-times daily prayers is used to measure a student's level of compliance. It demonstrates that religious education nowadays has a tendency toward school-based rather than home and social learning-oriented one. As a result, students internalize their religions more as knowledge, not as a way of life that necessarily can grow ethos and social ethics.

Such educational situation has also brought cultural shocks to students when they come to play roles in societies where religious pluralism is inevitable. Religious prescriptions they received in schools turned to be far from social realities wherein there is also a gap between social and personal morals. In addition, since religion is taught in schools as 'ready-products', not as a method of understanding religion itself, students may find those products no longer actual in realities and, therefore, difficult to implement. They may also see their religious knowledge as alien and a gift from others because it did not stem from their own comprehension and their religious conscience. This sense of self-disconnected usually happens to those who are grown up but emotionally immature and lack of religious intellectuality (Abdurrahman, 1995: 231). Scribner (2007, 30) said that in Japan, all educational practices should put an emphasis on the component of moral education. Likewise, Islamic education should be more focused to teach values and moral components so that the students are able to well behave in life.

For this purpose, there is a need for a more comprehensive Islamic educational system, which does not merely put a serious attention to the cognitive aspect of religious apprehension, but also seriously take into account the affective and religiosity sides. The practices of religious education should not confine themselves into the delivery of teaching materials to the students; rather, they should also take on the process of internalization initiated by the school through a series of activities that can implant religiosity upon the students. To achieve this goal, it is necessary to bring forth a hidden curriculum worked out by all the parties in a given school.

It is not an exaggeration to say that the current generation is facing the most critical situation. This is because they now live in the most unstable era where societies have been developing so

rapidly and radically, from agrarian-traditional to industrial as in modern cities. This radical change has also created a sense of crisis among the young people who are at the stage of self-defining. Weak foundations of families, schools as a guinea pig, fragile and unfriendly societies with limited roles and job opportunities are among the factors that contributed to this condition. Consequently, the youngster become increasingly marginalized leading to self-inferiority and alienation (Abdurrahman, 1995: 227). It is in this situation that religious guidance finds its significance to help them identify the basic problems they are faced with.

Young people should be exposed to any possible opportunities after which they should make themselves well-prepared with the help of high mentality and working ethos, strong spirits and morality, as well as adequate competence. To successfully integrate all these three aspects in coping with the problems of the youth, religion should have strong basis and deep insights (Abdurrahman, 1995: 233). This effort requires a proper evaluation on the implementation of Islamic education such as in *madrasahs* and Islamic boarding schools (*pondok pesantren*) because it is only by doing so that improvements will be possible.

There are three major components in learning process, i.e. learning objectives, learning strategies, and evaluation or assessment (Hajarah, 2004: 3). These three components are supporting each other to enhance the good quality of learning process among students. Evaluation is an integral part of learning activities by which learning process will be well-guided.

The scopes of evaluation works include the assessment of students' learning outputs in the cognitive, affective, and psychomotor aspects in correspondence with the expected competence. The cognitive aspect is measured through tests, the affective through questionnaires, interviews, as well as observation. The psychomotor aspect is assessed through observation.

The current practices of evaluation in education are inclined toward the cognitive aspect, giving less attention, if any, to the affective one. This is because the evaluation models of the affective aspects have not been yet developed in all subjects, including Islamic Education which is in nature rich of affective values.

The research data on multi-intelligence showed that emotional quotient or affective capacity contributes to one's success both in professional works and life by around 80%, psychomotor by only 15%, and cognitive by only 5%.

The dominance of cognitive aspect is found also in the educational practices in Indonesia as reflected in the learning process and assessment. This cognitive aspect is mainly reflected in the four subjects, i.e. Language, Math, science, and social science. On the other hand, the affective is translated into the subjects of religious education and civic education (Depdiknas, 2003: 27). Therefore, there is an urgent need for the development of evaluation models of the affective aspect in Islamic education.

Based on the background above, this study attempts to answer the following research questions: (1) How the evaluation models of the affective aspect are developed for Islamic education in *madrasahs* and Islamic Boarding Schools (*pesantren*)? the answers are expected to provide proper and comprehensive information in terms of the scopes, contents, and the practicality as well as the benefits for Islamic education; (2) How is the validity of the instruments used in the evaluation models of the affective aspect of Islamic education in the two Islamic education institutions?

II. RESEARCH METHODOLOGY

Models of development in this research refer to spiral model as put forward by Cennamo & Kalk (2005: 6). According to this spiral model, development includes defining the products aimed to be developed, designing the products, demonstrating, developing, and delivery. Based on this model being adjusted to Provus' evaluation model of gap focusing on inputs, processes, and outputs, the stages of the research and development in this paper include preliminary investigation and development design. The next steps are demonstration, test, evaluation, and revision of the design, development, and delivery.

In general, the stages of development above can be mapped out into three: (1) pre-development; (2) development; and (3) model application. The pre-development stage of evaluation model of the affective aspect in Islamic education is preliminary studies, including theories, relevant literary reviews, and field observations. The development stage of the model includes activities of defining and designing the evaluation models of the affective aspect, as well

as delivering the model. The implementation stage of the evaluation model of the affective aspect in Islamic Education includes testing, evaluation and revision, validation, and delivery of the model.

The procedures of the evaluation of the affective aspect in Islamic Education being developed consist of inputs, process, and output (IPO). The instruments of the evaluation models being tested were the ones aimed to assess the input, the process, and the output. The tests were conducted to examine two aspects: clarity (understandability) of the instruments and the compatibility of the measurement model, i.e. whether or not the conceptual model are theoretically in accordance with the empirical data. The tests on the clarity of the instruments were carried out on the input, process, and output of the Islamic Education. The clarity of the instruments was determined by the clear guidance of the instruments, their scopes, languages, and syntaxes. The clarity of the scopes includes clear indicators that underlie the formulation of the instruments. The language aspect was assessed by examining whether the propositions are communicative and the sentences are easy to grasp. The syntaxes were assessed on the basis of font type and size as well as format of the instruments.

III. RESEARCH FINDINGS AND DISCUSSION

The index validity of the input construct, which includes teachers' competence, curriculum, facilities and infrastructures, is shown in Table 1.

Table 1. Component of Input Evaluation

Dimension	Indicator	S	iv	z	p
Teacher's competence	1. Apprehending the students very well	21	1.000	3.381	0.000
	2. Able to design the learning process	21	1.000	3.381	0.000
	3. Able to implement the learning design	21	1.000	3.381	0.000
	4. Able to conduct the assessment	19	0.905	2.704	0.003
	5. Mastering the learning materials	21	1.000	3.381	0.000
	6. Able to relate the topic of the subject to other relevant aspects	19	0.905	2.704	0.003
	7. Understanding contemporary issues	21	1.000	3.381	0.000
	8. Behaving consistently	19	0.905	2.704	0.003
	9. Being wise and dignified	21	1.000	3.381	0.000
	10. Able to be a role-model	19	0.905	2.704	0.003
	11. Able to communicate effectively	21	1.000	3.381	0.000
	12. Able to get along easily	17	0.810	2.028	0.021
	Average of Teacher's Competence	18.519	0.882	2.542	0.006
Facilities and Infrastructures	1. Codes of conduct	18	0.857	2.366	0.009
	2. School regulations	19	0.905	2.704	0.003
	3. Religious activities in school	20	0.952	3.043	0.001
	4. Laboratory for Islamic education	21	1.000	3.381	0.000
	5. Mosque	21	1.000	3.381	0.000
	Average of Facilities and Infrastructures	18.167	0.865	2.423	0.008
Curriculum	1. Habituating good conducts	19	0.905	2.704	0.003
	2. Avoiding bad manners	19	0.905	2.704	0.003
	3. Understanding the stories about devout people	17	0.810	2.028	0.021
	4. Understanding human tasks and obligations	17	0.810	2.028	0.021
	5. Keeping one's self from bad characters	17	0.810	2.028	0.021
	6. Understanding the stories disobedient people	14	0.667	1.014	0.155
	7. Understanding the Qur'anic verses on religiously allowed, healthy, and nutritious foods, as well as the danger of alcoholic drinks	19	0.905	2.704	0.003
	8. Understanding the Qur'anic verses on faithfulness to Allah and His Messenger	21	1.000	3.381	0.000
	9. Understanding the Qur'anic verses on the greatness and power of Allah	19	0.905	2.704	0.003
	10. Understanding the Qur'anic verses on the grace of Allah and how to be thankful	19	0.905	2.704	0.003

11. Understanding the Qur'anic verses on simple ways of life and the order to help the poor	19	0.905	2.704	0.003
12. Understanding the Qur'anic verses on the utilization of natural resources	20	0.952	3.043	0.001
13. Understanding the Qur'anic verses on competition and goodness	21	1.000	3.381	0.000
14. Understanding the Qur'anic verses on doing good and avoiding evil (<i>amar ma'ruf nahi munkar</i>)	19	0.905	2.704	0.003
15. Understanding the Qur'anic verses divine tests	21	1.000	3.381	0.000
16. Understanding the Qur'anic verses on human responsibility in families and societies	21	1.000	3.381	0.000
17. Understanding the Qur'anic verses on the obligation of religious proselytizing	21	1.000	3.381	0.000
18. Understanding the Qur'anic verses on leadership	21	1.000	3.381	0.000
19. Understanding the Qur'anic verses on conflict resolution	19	0.905	2.704	0.003
20. Understanding the Qur'anic verses on fairness and honesty	19	0.905	2.704	0.003
21. Understanding the Qur'anic verses on personal, familial, and social character building	21	1.000	3.381	0.000
22. Understanding the Qur'anic verses on tolerance and social ethics	21	1.000	3.381	0.000
23. Understanding the Qur'anic verses on ethics in a meeting	19	0.905	2.704	0.003
24. Understanding the Qur'anic verses on working ethos	19	0.905	2.704	0.003
25. Understanding the Qur'anic verses on religiously allowed and good foods	21	1.000	3.381	0.000
26. Understanding the Qur'anic verses on the potentiality of the intellect and science	21	1.000	3.381	0.000
27. Understanding the Qur'anic verses on science and technology.	19	0.905	2.704	0.003
28. Understanding the hadiths on faithfulness to Allah and His Messenger	21	1.000	3.381	0.000
29. Understanding the hadiths on the greatness and power of Allah	21	1.000	3.381	0.000
30. Understanding the hadiths on the grace of Allah and how to be thankful	19	0.905	2.704	0.003
31. Understanding the hadiths on simple ways of life and the order to help the poor	18	0.857	2.366	0.009
32. Understanding the hadiths on the utilization of natural resources	21	1.000	3.381	0.000
33. Understanding the hadiths on doing good and avoiding evil (<i>amar ma'ruf nahi munkar</i>)	19	0.905	2.704	0.003
34. Understanding the hadiths on divine tests	21	1.000	3.381	0.000
35. Understanding the hadiths on the obligation of religious proselytizing	21	1.000	3.381	0.000
36. Understanding the hadiths verses on human responsibility in families and societies	21	1.000	3.381	0.000
37. Understanding the hadiths on leadership	21	1.000	3.381	0.000
38. Understanding the hadiths on conflict resolution	21	1.000	3.381	0.000
39. Understanding the hadiths on fairness and honesty	21	1.000	3.381	0.000
40. Understanding the hadiths on tolerance and social ethics	19	0.905	2.704	0.003
41. Understanding the hadiths on ethics in a meeting	21	1.000	3.381	0.000
42. Understanding the hadiths on working ethos	21	1.000	3.381	0.000
43. Understanding the hadiths on religiously allowed and good foods	19	0.905	2.704	0.003
44. Understanding the hadiths on the potentiality of the intellect and science	19	0.905	2.704	0.003
45. Understanding what is going to happen after the end of the world	20	0.952	3.043	0.001
Average of Curriculum	18.321	0.872	2.475	0.007
Average of Input Component Construct	19.774	0.942	2.966	0.002

Table 1 shows that the content validity of the input component construct is 0,942 with a probability of $p = 0,002$, which means "good".

Table 2. Evaluation Process Component

Dimension	Indicator	S	Iv	z	p
Planning	1. Syllabi	17	0.810	2.028	0.021
	2. RPP	21	1.000	3.381	0.000
	Average of Planning	19.083	0.909	2.733	0.003
Implementation	1. Implementation of Learning	21	1.000	3.381	0.000
	Average of Implementation	17.912	0.853	2.337	0.010
Assessment	1. Affective Model of assessment	21	1.000	3.381	0.000
	2. Methods of Behavior Assessment	21	1.000	3.381	0.000
	3. Methods of Competence Assessment	18	0.857	2.366	0.009
	Average of Assessment	18.946	0.902	2.686	0.004
	Average of Evaluation Process Component	19.833	0.944	2.986	0.001

Table 2 shows that the content validity of the construct of evaluation process component is 0,944 with a probability of $p = 0,001$, which means "good".

Table 3. Table 3. Output Evaluation Component

Dimension	Indicator	S	Iv	z	p
Discipline	1. Having a self-discipline in following the courses	21	1.000	3.381	0.000
	2. Abiding to the rules	21	1.000	3.381	0.000
	3. Having an orderly manner in classes	21	1.000	3.381	0.000
	4. Working on assignment properly	19	0.905	2.704	0.003
	5. Being obedient to the teachers' commands	17	0.810	2.028	0.021
	Average of Discipline	18.6	0.886	2.569	0.005
Respect	1. Greeting	21	1.000	3.381	0.000
	2. Willing to listen others	21	1.000	3.381	0.000
	3. Thankful to gifts	21	1.000	3.381	0.000
	4. Respecting parents	19	0.905	2.704	0.003
	5. Using good words	19	0.905	2.704	0.003
	6. Being polite	21	1.000	3.381	0.000
	7. Appraising friends	20	0.952	3.043	0.001
	Average of Respect	17	0.810	2.028	0.021
Responsibility	1. Getting done the assignments properly	21	1.000	3.381	0.000
	2. Being consequent with utterance	21	1.000	3.381	0.000
	3. Following rules	19	0.905	2.704	0.003
	4. Considering the consequence of acts	19	0.905	2.704	0.003
	5. Getting used to orderly and organized manners	17	0.810	2.028	0.021
	Average of Responsibility	18.4	0.876	2.502	0.006
Religious Observance	1. Praying five times a day	21	1.000	3.381	0.000
	2. Saying prayers in and after the shalat	17	0.810	2.028	0.021
	3. Reading the Qur'an	21	1.000	3.381	0.000
	4. Fasting during Ramadhan	21	1.000	3.381	0.000
	5. Wearing Islamic dresses	19	0.905	2.704	0.003
	6. Eating and drinking in an Islamic way	21	1.000	3.381	0.000
	7. Socializing in an Islamic way	19	0.905	2.704	0.003
	8. Sacrificing on behalf of Islam	19	0.905	2.704	0.003
	9. Attending Islamic activities	21	1.000	3.381	0.000
	10. Helping the needy people	19	0.905	2.704	0.003
	Average of Religious Observance	18.3	0.871	2.468	0.007
	Average of Output Component Construct	20	0.945	2.992	0.001

Table 3 shows that the content validity of the output evaluation component is 0,945 with a probability of $p = 0,001$, which means "good".

IV. CONCLUSION AND SUGGESTION

The evaluation model in the affective domain in Islamic education instruments overall that has good content validity index. Thus the evaluation model in the affective domain in Islamic education instruments can be done to test for testing the model. It is recommended for expert justification should be no shared understanding of the instrument of accession of the affective domain in Islamic education, and that the score be obtain balanced scoring.

REFERENCES

- [1] Abdurrahman, M. (1995) Islam transformatif. Jakarta: Pustaka Firdaus.
- [2] Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. Educational and psychological measurement., 40(4), p. 955-959.
- [3] Cennamo, K. & Kalk, D. (2005). *Real world instructional design*. Thomson. Wadsworth.
- [4] Depdiknas. (2004). *Penilaian Kelas*. Jakarta: Pusat Kurikulum Balitbang Depdiknas, p.27
- [5] Hajaroh, M. (2004). *Pengembangan evaluasi afektif mata kuliah pendidikan agama Islam di prodi D-II PGSD guru kelas Universitas Negeri Yogyakarta*. Penelitian DIP UNY, p.3
- [6] Provus, M. M. (1969). *The Discrepancy evaluation model: an approach to local program improvement and development*. Pittsburg Public Schools.
- [7] Scribner, S. (2007). Civic and moral education in japanese schools, *Leadership for Student Activities*, p.30

Developing Science Process Skill Instrument of Islamic Senior High Schools

Kadir ¹ Sri Wahyuningsih ², Abd. Rahman A. Ghani ³

¹ kadir@uinjkt.ac.id, Dosen UIN Syarif Hidayatullan Jakarta

² umifathur.18@gmail.com, Guru Biologi MAN 17 Jakarta

³ abdulrahman.ghani@yahoo.co.id, Direktur dan Dosen SPs. UHAMKA

Abstract - The objective of this study was to develop a reliable and validity instrument for students' science process skill in biology subject at Islamic Senior High Schools in DKI Jakarta. Five Islamic Senior High Schools in DKI Jakarta was selected by random sampling technique from 22 Islamic Senior High Schools in DKI Jakarta which focused on twelfth grade. This study used a research and development method. The result of study reveals that (1) From 48 items representing every dimension and the indicator showed that only 46 items are categorized valid after assessed by 20 panelist based on the value of CVR, which obtained the highest CVR value of 1.00 and the lowest was 0.30. (2) From the 46 items which had been tested and showed that only 41 items were categorized as valid based on item analysis 'ITEMAN' program, it was founded that items were valid which have pointbiserial coefficient > 0.20 and reliability coefficient of 0.867 which means that the reliability of the instrument science process skills are very good. (3) From the 41 items assessed for compliance which is valid between the instrument models and the data in the field using CFA, with construct reliability 0.986 and obtained minimum value Function Fit Chi-Square = 744.06 ($P = 0.06$) which indicates that the instrument developed a new appropriate approaches. Based on the result above, it can be drawn conclusion that *observing, classifying, measuring, inferring, predicting, and communicating* can measure the students' science process skill.

Keywords: *validity, reliability, science process skill, confirmatory factor analysis.*

I. INTRODUCTION

Achieving of the Indonesian children in TIMSS 2011 on the rank 40 and at the position 38 from 42 countries shows the low of learning achievement in learning science and mathematics in our country. Instruction or learning has not ... between hands-on and minds-on, which is not give any impact and benefit to the science teachers because testing system that only measuring the concept mastery and definitions. (Rustaman, 2007: 818). Another resemble opinion as stated by Elin Driana (Kompas, December 12, 2012) that "instruction in top countries rank in PISA is more focuseing on higher reasoning level has changed instruction which material mastery oriented for preparation to test and memorizing and drills. The changing of focus on instruction needs the teachers who have ability in creating learning atmosphere which supporting it"

The main cause of the low science process skill is lack of thinking skill because of the lack of and less of tests which measure the achievement about thinking. process skill involves intellectual skill, manual, and social. Intellectual skill is process skill which involving thinking, manual skill involves using tools and equipments, and social skill involves social interaction in teaching and learning and discussing based on the result of observation.

Rustaman (2007) states that science process skill is scientific inquiry, procedure that guides to gain knowledge and to give definition which more meaningful to the students. Through developing science process skill can be given the opportunity to develop the concept and process simultaneously. According to Rezba, et al. (Patta Bundu, 2006: 24), there are some types of science process skill, namely: observing, classifying, measuring, inferring, predicting, and communicating.

Based on the problems above, it indicates that lack of empowering the test property which measuring the achievement about thinking as one of the cause of low science process skill in science education. So that it is needed an instrument which objective, valid, reliable to assess students' science skill. Based on some problems elucidated above, the objective of the research are to develop science process skill in Islamic Senior high school in Biology subject.

The results of this research are expected to give some significant contributions not only theoretically but also practically as follows: (1) Academically, can enrich the measurement theory, especially science process skill; (2) Practically, can be used to analyze a profile of mastery of science process skill for students.

II. METHOD

This research method used in this investigation was *Research and Development*. Next, this research adapted Borg & Gall (Sugiyono, 2010: 271) which are consist of 10 stages to be 5 stages, namely: (1) product analysis which will be developed; (2) developing of initial product; (3) product validation; (4) field try-out; and (5) product revision.

There are some stages in developing science process skill instrument, they are as follows:

1. Developing blue print.
2. Constructing item through discussion among Biology teachers.
3. Revising instrument based on expert judgement.
4. Trying-out stage one to know the essential indicator and readability.
5. Trying-out empirically to know validity and reliability.
6. Testing *Fit Model* through conformatory factor analysis.

A. Validity Testing

In developing instrument to measure the science process skill, the content validity, theoretically conducted by an expert and 20 Biology teachers to make sure that the test items are relevant and represent all domain are measured. Content validity is measured by using *Content Validity Ratio* (CVR) with Lawshe formula as follows:

$$CVR = \frac{\left(Ne - \frac{N}{2} \right)}{\frac{N}{2}}$$

Keterangan :

Ne: total number of pannelist who assess esensial, N: total of panelist

Valid item if $CVR \geq$ minimum value CVR for 20 raters as 0.42. empirical stage one with ITEMAN program. Coefficient validity stated as a good category if coefficient is higher than 0.20.

Furthermore conducted parameter estimated testing on loading factor (λ) by using CFA. The criteria which using was valid items if $\lambda > 0.5$ or t-statistical testing of parameter λ higher than 1.96.

B. Reliabilitas

The concept of reliability in this respect meant that reliability of measurement tool related to the problem of measurement error. According to Menurut Sudaryono (2012: 155), reliability related to the error in choosing sample which refers to inconsistency of the measurement result if the measurement reconducted on different group.

The level of reliability in this research estimating based on the score of reliability coefficient calculated by using formula (Wijanto, 2008: 175):

$$\text{Construct reliability}(CR) = \frac{(\sum \text{loading})^2}{(\sum \text{loading})^2 + \sum e_j'}$$

$$\text{Variance extracted (VE)} = \frac{\sum \text{loading}^2}{N}$$

Model reliability estimation regarded good if $CR \geq 0,70$ dan $VE \geq 0,50$.

III. RESULT

The findings of each stage of instrument development of science process skill is presented as follows.

A. Theoretical Validation (Tried-out stage 1)

Tried-out rationally by expert on the property of instrument of science process skill was conducted to expert teacher and some Biology teachers and so as expert teacher. The panelists perused each of the basic science process skill based on the indicators which were designed, and confirmed the fitness of test items with the way of giving score for essential items (E) and inessential test items (IE). The results of grading given by the panelists on test items of science process skill, show that there are 6 dimensions which are consist of 48 items have been validated and there were 46 items considered valid and 2 items not valid, with the scale score of CVR from 0.30 up to 1.00.

B. Empirical validation (The Second Stage Try-out)

The result of empirical testing of the instrument of students' science process skill to 397 grade XII of Islamic senior high school in DKI Jakarta. Meanwhile, the results based on ITEMAN program, reveals that there are 5 test item were not valid (score of Point Biserial < 0.20) they are number 1, 5, 8, 10, and 26. So, the total number of valid items are 41 with reliability coefficient (KR-20) is 0.867, this value or score can be inferred that the test items which are developed have high reliability.

C. Confirmatory Factor Analysis (CFA)

The result of parameter analysis λ (loading factors) used CFA reveals that all item has value or score $\lambda > 0.5$ or statistical value/score based on t-test > 1.96, with scale score λ : observation dimension consist of 13 test item, was (0.75–2.75), classification dimension consists of 5 items (0.56–2.17), measure 2 items (1.00), inference 5 items (1.00–1.13), prediction 4 items (0.08–1.00), and communication dimension consist of 5 items (0.52–1.00). This results show that there are 41 items have good validity to measure 6 dimension of science process skill.

Next, the result of Construct Reliability analysis (CR) and variance extracted (VE), shows that observation (CR=0.99; VE= 0.93), classification dimension (CR=0.97; VE= 0.88), measure (CR=1.00; VE= 1.00), inference (CR=0.98; VE= 0.92), predict (CR=0.87; VE= 0.77), and communication dimension (CR=0.87; VE= 0.78). The reliability testing shows that the score CR > 0.70 and VE > 0.50. So that the reliability estimation both all dimension and individual one based on dimension and factor shows that the instrument of science process skill has a very good internal consistency.

The result as depicted through the path diagram below shows that the instrument of science process skill based on dimensions, covers classification observation, measure, inference, predict, and communicate (Figure 2).

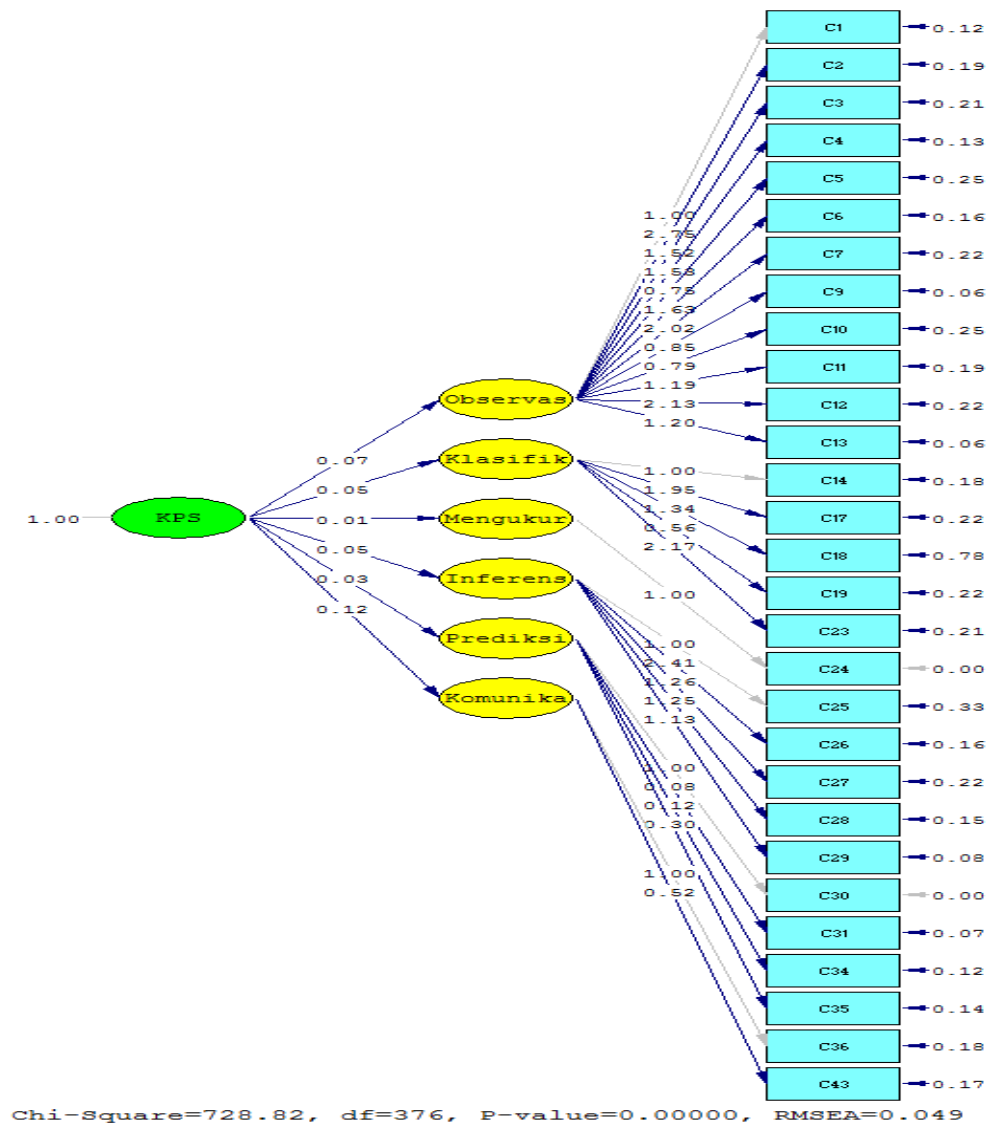


Figure 1. Path Diagram of the estimation result with Second Order CFA

Furthermore the testing of the model appropriateness meant to measure about to what extent is the measurement model which is proposed fit or appropriate with the research data. The following statistical testing is to test which is the fit/appropriate indicator. The statistical formula used were Chi-Square, RMSEA, ECVI, AIC, CAIC, NFI, NNFI, PNFI, CFI, IFI, RFI, GFI, CN, RMR, dan AGFI. The result based on analysis through using Fit Model can be presented in following table.

Table 1. Summary of fit model

GOF Criteria	Fit Indikator	Estimation	Assessment
<i>Chi-Square</i> (p)	Small score $p > 0.05$	$\chi^2 = 744.06$ ($p = 0.06$)	good
NCP Interval	Small interval score that narrow	352.82 (280.18 ; 433.26)	Fairly good
RMSEA	$RMSEA \leq 0.08$ $P \geq 0.05$	0.049 $P = 0.65$	good fit
ECVI	Small score and near to ECVI <i>Saturated</i>	$M^* = 2.14$ $S^* = 2.20$ $I^* = 2.98$	good fit
AIC	Small score and close to AIC <i>Saturated</i>	$M^* = 846.82$ $S^* = 870.00$ $I^* = 1323.11$	good fit
CAIC	Small score and close to CAIC <i>Saturated</i>	$M^* = 1140.87$ $S^* = 3038.01$ $I^* = 1323.11$	(good fit)
NFI	$NFI \geq 0.90$	0.34	Fairly good
NNFI	$NNFI \geq 0.90$	0.44	Fairly good
CFI	$CFI \geq 0.90$	0.48	Fairly good
IFI	$IFI \geq 0.90$	0.51	Fairly good
RFI	$RFI \geq 0.90$	0.28	Fairly good
CN	$CN \geq 200$	236.62	(good fit)
RMR	Stand RMR ≤ 0.05	0.012	Fairly good
GFI	$GFI \geq 0.90$	0.89	Marginal fit
AGFI	$AGFI \geq 0.90$	0.87	Marginal fit

M^* = Model; S^* = Saturated; I^* = Independence

Based on the result as seen in the table above, the main fit indicator has the *Goodness of Fit* and it is categorized 'good'. This means that the result of conceptual model of science process skill testing which is peoposed is 'fit' or appropriate with the data.

IV. DISCUSSION

The instrument of students' science process skill of Islamic senior high school in Biology subject which is developed based on theoretical analysis and expert investigation then followed by trying-out empirically that has been developed can be used for measuring science process skill.

Research finding reveals that, observed from the average factor load, basic science process skill, in the sequence of the lowest-highest validity the skill in succession, namely: (a) predict, (b) communicate, (c) inference, (d) classify, (e) observe, and (f) measure. Inference skill is the skill to draw conclusion and description from the observation result. If the observation is an experience which is achieved through one or two human senses, so inference is interpretation or description on the observation result. This skill is the most important science process because the correctness of the science or knowledge achieved is depend on appropriateness and accurateness of the observation result. The ability to conduct an observation the ability to think is the basis skill in science, and including the skill dimension which is categorized as a hard aspect in this research. The finding of this research in line with Subadi's research result (2009), which reported that the ability to think divergent in respect of data/information recording skill is the basic skill which is the most difficult, meanwhile the easiest is observation skill. The ability to think divergence in the respect of inference making skill and process skill is the most difficult skill, while the easiest skill is making prediction. Divergent thinking ability in doing investigation shows the lowest score up to the highest one in designing, doing, reporting, and investigating.

The finding of this research support the theory and concept of science process skill which was founded by Rustaman (2007) that the science process skill was the science inquiry domain, procedure which guides in achieving knowledge and giving definition that is more meaningful for students. Furthermore Rezba, et al. (1995) gives more detail description between basic science process skill and integrated science process skill with the six basic science process skill, which covers: observing, classifying, measuring, inferring, predicting, and communicating.

Referring to the result of model testing, both wholeness and individually and learning number of loading factor and construct reliability and the result of testing fit model and relevant findings

can be summed up that the model of instrument measurement which is developed can be accepted as the assessor of instrument for science process skill in Islamic Senior high School, especially in Biology subject.

V. CONCLUSIONS

Based on finding and discussion of the research results, it can be drawn some conclusions as follows.

1. Developing instrument for science process skill for Islamic Senior high School, especially in Biology subject Pengembangan instrument keterampilan proses sains siswa di Madrasah Aliyah dalam mata pelajaran Biologi conducted through 5 stages, namely (a) product analysis which is developed which covers: theoretical analysis and discuss the previous related research result; (b) developing prior product; (c) product validity, covers the content validity through expert judgement and panneltist; (d) field try-out, consist of empirically tested on the first stage and second stage with CFA technique; and (e) product revision, including improvement post-try-out by paying attention to the expert consideration.
2. Dimension or factor which basis the science process skill of the Islamic Senior high School, especially in Biology subject achieved as many 6 dimension, namely: observing, classifying, measuring, inferring, predicting, and communicating which are measured by 41 items which have excellent validity.
3. Empirical validity on the item of science proces skill which is on excellent category. So that, it can be used in teaching and learning Biology subject.
4. The instrument of science process skill for Islamic Senior high School, especially in Biology subject has composite reliability as 0.986 or excellent, which comprises of doing observing 0,99, classifying 0.97, measuring 1.00, inferring 0.98, predicting 0.87, and communicating skill 0.87. Besides that, the result of conceptual model testing science process skill which is proposed is fit with the data.

A. Suggestions

Based on the reseacrh conclusion, it can be delivered some suggestions as follows:

1. The indicator of science process skill on the blue print of instrument need to be developed more variety. This respect can be rearrange the relevant items.
2. Developing the instrument need to be expanded its population not only to scope of area but also the level of education.
3. Considering that science process skill has an important role in learing successfulness, especially oin Biology subject, so it is suggested to the teachers in conducting the teaching and assessment with problem based learning dan project based learning approach.
4. Sciences and mathematics teacher should be given about science process training and constructing higher order thinking test (HOT).

REFERENCES

- [1] Brown, G., *Assessing Student Learning in Higher Education*, London: Routledge. 1996.
- [2] Bundu, P., *Penilaian Keterampilan Proses dan Sikap Ilmiah dalam Pembelajaran Sains Sekolah Dasar*, Jakarta: Departemen Pendidikan Nasional, 2006.
- [3] Driana, E., "Gawat Darurat Pendidikan," *Kompas*, 14 Desember 2012.
- [4] Lawshe, C. H.. *A quantitative approach to content validity*, By Personnel Psychology, INC, 1975.
- [5] PISA (Programme for International Student Assessment), 2012 Result, diakses dari (www.oecd.org), pada Desember 2014.
- [6] Rustaman, N.,. *Ilmu dan Aplikasi Pendidikan*. Bandung: Pedagogiana Press, 2007.
- [7] Subali, B., "Pengukuran Keterampilan Proses Sains Pola Divergen dalam Mata Pelajaran Biologi SMA di Provinsi DIY dan Jawa Tengah." Disertasi. UNY, 2009.
- [8] Sudaryono, *Dasar-dasar Evaluasi Pembelajaran*, Yogyakarta: Graha Ilmu, 2012

ONLINE EXAM MODEL OF ITEM RESPONSE THEORY BASED CAT USING MOODLE LEARNING MANAGEMENT SYSTEM

Khairawati

Abstract - This research aims to find a procedure to establish a valid question bank for the purposes of the Computerized Adaptive Test (CAT) on the e-learning system of the Faculty of Tarbiya and Teacher Training of the Institute of Islamic Studies (FTIK IAIN) Pontianak. The method used was Research and Development (R & D). Subjects were fifth semester students of FTIK IAIN in the academic year of 2015/2016 selected through purposive sampling, with 35 students being samples for phase 1 research, and 10 students for phase 2. The data were multiple choice questions test results using the P&P Test and CAT. The findings showed that creating a question bank for the purpose of CAT should be well planned because (1) the development of question bank is a very expensive investment; (2) it requires specific expertise; and (3) the construction to meet the Item Response Theory is very difficult. Results of analysis of the test results of the Subject of Research Methodology 1 showed that the item discrimination will produce more question items under *good* category when using CAT (40.00%) than the P&P Test method (28.58%).

Keywords: *Computerized Adaptive Test, P&P Test*

I. INTRODUCTION

Exam is one of the ways to evaluate the teaching and learning process. In the world of education, particularly in higher educational institutions, the exam is intended to measure the achievement of the goal of teaching courses to students as learners, so that students can determine the level of ability in understanding the subject being taken. If the result is not optimal, the learning process should be improved both in terms of quality and quantity.

Rapid development of the technology in communication and electronics has contributed to the education sector to enjoy an increase in quality, speed, convenience and ease. Conventional tests have also shifted toward computerization, one of which is the online exam. Even the entrance tests and graduation of an institution or profession are now using online exams. It can be seen from the development testing methods at this time, such as Computer Assisted Test (CAT) that is already commonly applied employee recruitment exams in many countries.

In Indonesia, the Ministry of State Apparatus Empowerment and Bureaucratic Reform requires all agencies to use computer-based test (CAT) for the selection of candidates for Civil Servants in 2014 (Tempo, 2014 Mandatory Computer Based Test for prospective civil servants, 2014). CAT is a test application conducted using a web-based computer for recruitment of employees. In addition to recruitment of employees, the National Exam (UN) in 2015 began to implement an online exam.

In light of the development of evaluation in education above, the researcher would like to develop a model of online exam of Computerized Adaptive Test (CAT) based on Item Response Theory (IRT) using the Learning Management System (LMS) Moodle software.

With the availability of online exam system, it is expected to help examiners to administer final exams (UAS) online, effectively and efficiently in the Faculty of Tarbiya and Teacher Training (FTIK) IAIN Pontianak.

A. LIETERATURE REVIEW

Computerized Adaptive Test or CAT was first applied in tests on intelligence by Binet in 1908 (Weiss, 1985). CAT in its development was also applied in the field of education (Lord, 1980). Lord examined CAT because test length that has been set was inefficient to measure the ability of the test taker. With the test package, participants with poor performance or highly capable test

takers could not be distinguished. Lord believed that the length of the test can be shortened without losing accuracy in the measurement. Test items should be given to each participant by selecting items that can provide maximum information about the ability of the test taker. In the theory of measurement each person taking the test is set in a group of items.

CAT can only be administered with a computer. Computers can store extensive test information (test items and characteristics = item bank) and display items that match the ability of the test taker. Computers can also be set up, and provide each participant's test scores on tests in accordance with their potential in a quick and accurate way (Bunderson, Inouye, & Olsen, 1989; WAINER, 1990).

A scheme of the Process of *adaptive testing* can be seen in Figure 1.

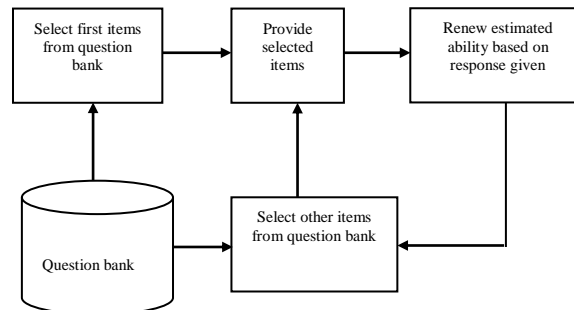


Figure 1. Process of Adaptive Testing

According to Hambleton, Zaal, & Pieters (1991), to date, research on adaptive test has focused on six areas: (1) selection of IRT models, (2) item bank, (3) starting point to test, (4) selection of the next test item, (5) assessment scoring, and (6) choice of method to decide when to end the test.

Item response theory (IRT), which is also known by the name of latent traits theory (LTT) was proposed as a solution to the weaknesses of the classical test theory (CTT). IRT has several advantages compared to CTT.

Embretson & Reise (2000) suggested 10 advantages of IRT over CTT, i.e.: (1) measurement standard deviation or standard error of measurement (SEM) has different score (or patterns of response), but it is common among the population; (2) the test is shorter and may be more reliable than a longer test; (3) comparison of test scores between various formats will be optimal if the difficulty level of the test varies between participants; (4) unbiased estimations can be obtained from the samples that are not representative; (5) test scores are significant when compared with item characteristics; (6) interval scale is achieved by using a more logical measurement model; (7) test with a mixture item format can produce optimal test scores; (8) scores that change can be compared significantly if the initial score is different; (9) results of the factor analysis on rough score data item can generate a full information factor analysis; and (10) properties of the item as a stimulus may directly relate to the psychometric properties.

There are three models of IRT (1) One parameter Model (Rasch Model), which is to analyze the data that focuses on parameters of the level of difficulty. (2) Two-parameter model, which is to analyze the data that focuses on the parameters and level of difficulty of item discrimination. (3) Three parameter Model, i.e. to analyze the data that focuses on the parameters about the difficulty level, item discrimination, and guessing (Hambleton, Swaminathan, & Rogers, 1991: 12-18).

One-parameter logistic model (1 PL) has the following formula:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

Where:

$P_i(\theta)$ = The probability of a subject that has the ability to answer the item to the θ -i correctly.

$i = 1, 2, 3, \dots, n$.

θ = level of ability of test taker

b_i = Parameter of level of difficulty of item to i

n = number of item in the test

e = transcendental number which is 2.718

Two-parameter logistic model (2 PL) can be written as follows:

$$P_i(\theta) = \frac{e^{D a_i(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

Where:

$P_i(\theta)$ = The probability of a subject that has the ability to answer the item to the θ -i correctly.

$i = 1, 2, 3, \dots, n$.

θ = level of ability of test taker

b_i = Parameter of level of difficulty of item to i

a_i = Index of item discrimination to-i

n = number of item in the test

e = transcendental number which is 2.718

D = scaling factor which is worth 1,7

Two-parameter logistic model (2 PL) can be written as follows:

$$P_i(\theta) = C_i + \frac{(1 - C_i) e^{D a_i(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

Where $B_i(\theta)$, b_i , a_i , and D are defined as two-parameter models. Additional parameter in the third model is C_i which is called pseudo-change level parameter.

MOODLE stands for Modular Object-Oriented Dynamic Learning Environment. Moodle is a learning management system software developed by using pedagogical principles. It is generally used for blended learning, distance education, e-learning projects in schools, universities, workplaces, and other sectors (Moodle, 2015).

According to Myrick (2010: 6) using Moodle as a system test will bring considerable changes to students and educators. Therefore, it is necessary to consider the following questions: (1) What will be tested, (2) items that will be used, (3) the ability of students to use computers and the internet, (4) experience in using Moodle, (5) online learning experience, and (6) question bank.

II. METHOD

This type of research is the development research, i.e. developing a model for measuring the achievement of students of FTIK IAIN Pontianak. The expected result of the development of exam system is a CAT program that can be applied to the final exam system of the FTIK IAIN Pontianak. The procedure for development consists of four stages: the first stage is the stage of identification of the subjects that will have CAT; the second stage is creating a question bank; the third stage is simulation of the workings of CAT, and the fourth stage is setting up CAT on the Final Exam system of FTIK IAIN Pontianak.

Steps of applying a learning model of online exam system based on Item Response Theory-based Computerized Adaptive Test using Moodle Learning Management System are as follows:

First, the entire question bank is incorporated into a learning environment IAIN Pontianak which is located on <http://vleiaainptk.ac.id/>. At this stage, the lecturers also direct and motivate the importance of online test models for the students through face to face meetings in the classroom.

Second, to be able to take the exams online in a virtual classroom, as well as working on the problems that have been provided through the web-based media, students must register online. Then they can sign into the virtual classroom that has been provided.

Third, if the students have finished answering on-screen questions, they can proceed to the next question.

Fourth, the questions that have been answered by the students will be corrected immediately by the computer, based on the answers to questions that have been incorporated into the virtual learning environment (VLE) of IAIN Pontianak.

In a limited experiment, the experimental method of single one shot case study was applied. In this experiment, 10 students are given treatment in the form of access to the VLE and answer

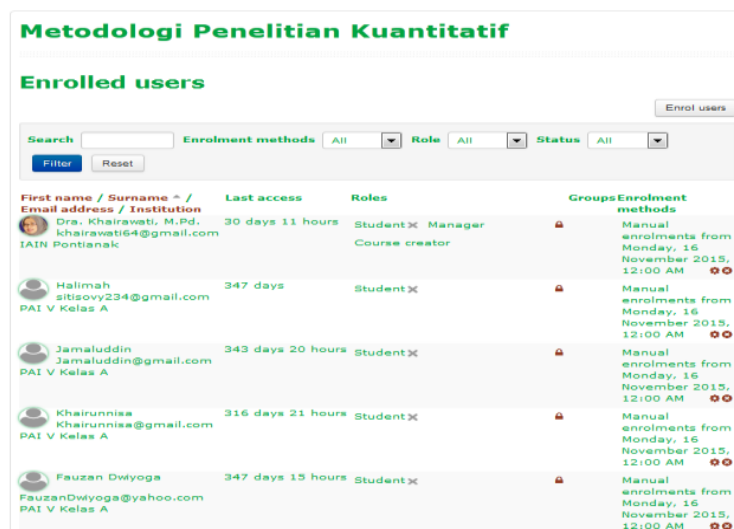
the questions that have been provided in the virtual classroom. After receiving treatment, 10 students will be observed.

III. RESULTS AND DISCUSSION

Creating a Question Bank for the purpose of CAT To create a question bank of multiple-choice type for the subject of Research Methodology 1, the following stages were conducted:

First, setting the goals; second, determining the scope of the Subject and Topics for Research Methodology 1; Third, determining the type of questions - in this study the type of questions is multiple choice; fourth, determining the way of collecting the questions --in this study the collection process is conducted through: preparing the content outline of question items for Research Methodology 1, writing / preparing the question items, conducting a review of items and experiment; fifth, determining Human Resources, at least there are three components of the personnel involved in the development of item bank for Research Methodology 1, i.e. a consultant who has competence in evaluation, a clerk responsible for collecting the items and a data entry clerk; and sixth, determining financing aspects.

Stages for creating the question bank development in this research had been done in a well-planned manner, it is given that some problems may arise in developing a question bank, as reported by Yahya Umar (1999) as follows: (1) Development of a question bank is a very expensive investment. (2) Development of a question bank requires a specialized expertise. (3) Construction of items that meet the Item Response Theory is very difficult. (4) In the items of achievement tests, the requirement for item response theory is difficult to meet. The stages of preparing the items at e-learning portal or into the virtual class belonging to IAIN Pontianak located at <http://vleiaipntk.ac.id/> are as follows: The first step the researcher did was to enroll students who took Research Methodology 1 into the virtual classroom of Quantitative Research Methodology



The screenshot shows a web interface titled 'Metodologi Penelitian Kuantitatif' with a section 'Enrolled users'. It includes a search bar, filters for 'Enrolment methods', 'Role', and 'Status', and a table of enrolled users.

First name / Surname ~ / Email address / Institution	Last access	Roles	Groups Enrolment methods
Dra. Khairawati, M.Pd. khairawati64@gmail.com IAIN Pontianak	30 days 11 hours	Student X Manager Course creator	Manual enrolments from Monday, 16 November 2015, 12:00 AM
Halimah sitisovy234@gmail.com PAI V Kelas A	347 days	Student X	Manual enrolments from Monday, 16 November 2015, 12:00 AM
Jamaluddin Jamaluddin@gmail.com PAI V Kelas A	343 days 20 hours	Student X	Manual enrolments from Monday, 16 November 2015, 12:00 AM
Khairunnisa Khairunnisa@gmail.com PAI V Kelas A	316 days 21 hours	Student X	Manual enrolments from Monday, 16 November 2015, 12:00 AM
Fauzan Dwiyoga FauzanDwiyoga@yahoo.com PAI V Kelas A	347 days 15 hours	Student X	Manual enrolments from Monday, 16 November 2015, 12:00 AM

Figure 2. Students enrolled in a Virtual Classroom

The second step is the process of preparing the question in a virtual learning environment. The first thing to do is choosing the type of questions that will be developed. The display of the choice of questions as shown in the following image.

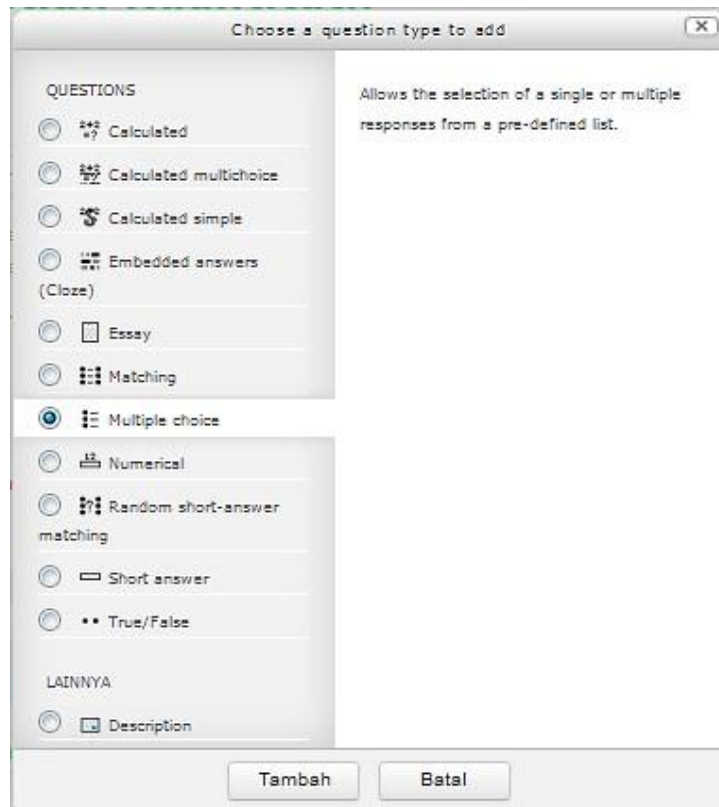


Figure 3. Selection of Types of Items in a Virtual Environment

After the process of making multiple choice questions is completed, then the end result will look like the following image.

Dalam merumuskan suatu judul penelitian, seorang peneliti harus merumuskan judul penelitiannya dengan baik. Berikut adalah daftar persyaratan suatu judul penelitian, kecuali:

Select one:

- ☐ a. Judul penelitian harus dapat menunjukkan permasalahan yang hendak diteliti
- ☐ b. Judul penelitian bukan suatu kalimat, oleh karenanya tidak perlu mempertimbangkan kaidah struktur kalimat
- ☐ c. Judul Penelitian harus dapat menunjukkan obyek penelitian dengan jelas dan kongkrit.
- ☐ d. Judul Penelitian mengacu pada referensi teoritikal

Figure 4. Multiple Choice Questions in CAT

The process of answering the first multiple choice question by some students is shown respectively in the following pictures

Question 1
Not yet answered
Marked out of 1,00

Penelitian yang dimaksudkan untuk mencari dan menunjukkan masalah beserta pemecahannya disebut:

Select one:

- ☐ a. Problem-Matter research
- ☐ b. Subject-Matter research
- ☐ c. Disciplinary research
- ☐ d. Problem-Solving research

Start again Save Fill in correct responses Submit and finish Close preview

Question 1
Correct
Marked out of 1,00

Penelitian yang dimaksudkan untuk mencari dan menunjukkan masalah beserta pemecahannya disebut:

Select one:

- ☐ a. Problem-Matter research
- ☐ b. Subject-Matter research
- ☐ c. Disciplinary research
- ☒ d. Problem-Solving research ✓

Your answer is correct.
The correct answer is: Problem-Solving research

Start again Save Fill in correct responses Submit and finish Close preview

Figure 5. Student 1 before and after answering question 1

Question 1
Not yet answered
Marked out of 1,00

Penelitian yang dimaksudkan untuk mencari dan menunjukkan masalah beserta pemecahannya disebut:

Select one:

- ☐ a. Disciplinary research
- ☐ b. Problem-Solving research
- ☐ c. Problem-Matter research
- ☐ d. Subject-Matter research

Start again Save Fill in correct responses Submit and finish Close preview

Question 1
Incorrect
Marked out of 1,00

Penelitian yang dimaksudkan untuk mencari dan menunjukkan masalah beserta pemecahannya disebut:

Select one:

- ☐ a. Disciplinary research
- ☐ b. Subject-Matter research
- ☐ c. Problem-Solving research
- ☒ d. Problem-Matter research ✗

Your answer is incorrect.
The correct answer is: Problem-Solving research

Start again Save Fill in correct responses Submit and finish Close preview

Figure 6. Student 2 before and after answering question 1

Question 1
Not yet answered
Marked out of 1,00

Penelitian yang dimaksudkan untuk mencari dan menunjukkan masalah beserta pemecahannya disebut:

Select one:

- ☐ a. Disciplinary research
- ☐ b. Problem-Solving research
- ☐ c. Problem-Matter research
- ☐ d. Subject-Matter research

Start again Save Fill in correct responses Submit and finish Close preview

Question 1
Incorrect
Marked out of 1,00

Penelitian yang dimaksudkan untuk mencari dan menunjukkan masalah beserta pemecahannya disebut:

Select one:

- ☐ a. Disciplinary research
- ☐ b. Subject-Matter research
- ☐ c. Problem-Solving research
- ☒ d. Problem-Matter research ✖

Your answer is incorrect.
The correct answer is: Problem-Solving research

Start again Save Fill in correct responses Submit and finish Close preview

Figure 7. Student 3 before and After answering question 1

In Figures 5, 6 and 7 it can be seen that the same question has an automatically different location or position of the answer choices for three students. This shows that random mode of answer choices in the online CAT system is functioning properly. It can also be used to anticipate or reduce the practice of cheating that often occurs among students.

A. Analysis of test results through the P & P Test and CAT

An analysis of the question items for the course of Research Methodology 1, fifth semester students of class A, Department of Islamic Religious Education (PAI) Faculty of Tarbiya and Teacher Training (FTIK) IAIN Pontianak in the academic year of 2015/2016 was carried out using a paper and pencil test (P & P Test). In this study, only two types of analysis were carried out i.e. the Item Discrimination Analysis and Difficulty Level Analysis.

Out of 70 items provided, the highest number of questions answered correctly by students was 47 or 67.14%, while the lowest was 38 or about 54.29%. The highest score obtained by the students was 67.21 while the lowest was 54.34. Meanwhile, out of 35 students who answered the 70 items, 31 students or 88.57% passed, while 4 or 11.43% did not.

Meanwhile, the analysis of the question items involved two parameters, i.e. the level of difficulty of the items and item discrimination. The results of the analysis of the items on the Research Methodology 1 course is as follows:

Table 1. Distribution of Item Parameter category using P&P Test

Level of difficulty (a)		
Description	Frequency	Test taker
Item Easy	25 Items	35.71%
Item Medium	23 Items	32.86%
Item Difficult	22 Items	31.43%
Item discrimination (b)		
Excellent	0 Item	0.00%
Good	20 Items	28.58%
Fair	29 Items	41.43%
Poor/needs revising	13 Items	18.58%
Discarded	8 Items	11.43%

The results of experiment using CAT showed that out of 10 items provided, the highest number of questions answered correctly by students WAS 8 or 80.00%, while the lowest was 3 or 30.00%. The highest score successfully obtained by the students was 80.00 while the lowest was 30.00. Meanwhile, out of 10 students answered 10 items, 6 or 60.00% of the students passed, while 4 or 40.00% did not.

Meanwhile, based on the item analysis, involving two parameters, i.e. level of difficulty and item discrimination on the test using the CAT, the results of the analysis of the course of Research Methodology 1 is as follows:

Table 2. Distribution of Item Parameter category using CAT Test

Level of difficulty (a)		
Description	Frequency	Test taker
Item Easy	3 Items	30.00%
Item Medium	6 Items	60.00%
Item Difficult	1 Item	10.00%
Daya Bada (b)		
Excellent	0 Item	0.00%
Good	4 Items	40.00%
Fair	3 Items	30.00%
Poor/needs revising	3 Items	30.00%
Discarded	0 Item	0.00%

Applying CAT in the e-learning system of FTIK IAIN Pontianak

In the Moodle e-learning applications that have been applied at IAIN Pontianak which is located at <http://vleiaipntk.ac.id>, there are facilities to create an exam or give a test online It is made possible because Moodle as a Learning Management System (LMS) which has a facility for developing a question bank as seen in the following picture:

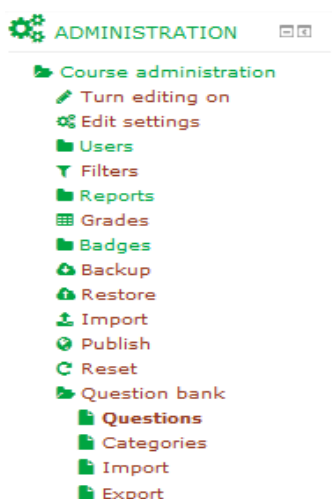


Figure 8. Menu for Creating Question Bank in LMS Moodle.

In the facility of preparing a question bank provided by Moodle, there are several types of questions, i.e.: true or false, matching, multiple choice, short-answer questions and making description.

Out of various types of item preparation facilities available in Moodle, this study used the type of multiple choice questions. Based on the experiment of preparing the exam online in the e-learning environment of IAIN Pontianak, test takers can get information such as when the test can be taken, how many minutes the test can be completed, which can be arranged through the timing facility. In addition, there is also a shuffle question facility, where the questions can be set to appear at random and shuffle within question facility, where the answer key of the same question can be randomized.

The next experiment was to apply online exam of Computerized Adaptive Test (CAT) based on Item Response Theory (IRT) using the Learning Management System (LMS) of Moodle at IAIN Pontianak. Several experiments were conducted to develop this model.

Unfortunately, the analysis facility available in the virtual learning environment using Moodle AT IAIN Pontianak has not supported the Item Response Theory. So the experiment of online exam of IRT-based CAT using Moodle learning management system could not be implemented through a virtual learning environment at IAIN Pontianak.

According to an IT consultant, Yusniardi, to develop CAT modules, we must also consider programming standards available in Moodle i.e. PHP - MySQL. Not all PHP and MySQL can be directly applied and integrated. We should create a special program with the PHP - MySQL based programming codes. To change the quiz modules already available in Moodle 1.9 into CAT modules, there are a few things to do, such as changing the quiz module database by adding several new fields on several tables. But unfortunately, to add PHP - MySQL based programming codes in the virtual learning environment IAIN Pontianak is currently not possible because the website hosting at <http://vleiaipntk.ac.id> is under the status of being rented. Hosting is a place to collect data required by a website and making it accessible via the Internet. Data here can be files, images, emails, applications / programs / scripts and database.

IV. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

Based on the research findings described in the previous section, the following conclusions can be drawn: *First*, the steps to developing a Question Bank for CAT purposes are as follows: (1) determining the purpose; (2) determining the scope of the subject and topics; (3) determining the type of questions; (4) determining how to collect questions, (5) determining the human resources; (6) determining the financing aspects. The stages in the development of question bank should be well planned because: (1) the development of a question bank is a very expensive investment; (2) it requires special expertise; and (3) the construction of items to meet the Item Response Theory is very difficult.

Second, the results of analysis using P & P Test showed that the question items for the subject of Research Methodology 1, based on the difficulty level, contained 35.71% of easy questions, 32.86% medium, and 31.43% difficult questions, while based on item discrimination there were no items that fell into the category of excellent, 28.58% into good category, 41.43% fair, 18.58% poor category so it should be revised, and 11.43% of the items should be discarded. Meanwhile, the results of analysis using CAT showed that the question items for the subjects of Research Methodology 1, based on the difficulty level, contained 30.00% of easy questions, 60.00% fair questions, and 10.00% difficult questions, while based on item discrimination, there were no items that fell into the category of excellent, 40.00% into good category, 30.00% fair, 30.00% about the ugly poor and should be revised, but no items should be discarded.

Third, the application of IRT-based CAT in the e-learning system of FTIK IAIN Pontianak could not be done because the Moodle of IAIN Pontianak (<http://vleiaipntk.ac.id>) did not support the application of Item Response Theory, so additional specific program that supports the implementation of an online exam is required.

B. Suggestions

First, it is necessary to develop a question bank for each course so that preparing tests will be easier, faster and more efficient. With the availability of question bank, the quality of test items

can be improved and guaranteed. In addition, it also serves as a form of an introduction to item response theory and the use of extensive computer knowledge at IAIN Pontianak.

Second, in the future it is necessary to conduct specific research to analyze questions items by applying three kinds of models of IRT, namely one-parameter logistic model (1 PL), two-parameter logistic model (2 PL) and three-parameter logistic model (3 PL) by utilizing IRT software which is available for free.

Third, it is necessary to provide a website hosting independently at IAIN Pontianak, so that the use of Information and Communications Technology to support lectures, research and community service will be more optimal.

REFERENCES

- [1] Hambleton, R. K., Zaal, J. N., & Pieters, J. M. (1991). Computerized Adaptive Testing: Theory, Application, and Standards. In R. K. Hambleton, & J. N. Zaal, *Advances in Educational and psychological Testing: Theory and applications* (pp. 341-366). Boston: Kluwer.
- [2] Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New York: Lawrence Erlbaum.
- [3] Moodle. (2015, July 23). *About Moodle*. Retrieved from Moodle: https://docs.moodle.org/29/en/About_Moodle
- [4] Myrick, J. (2010). *Moodle 1.9 Testing and Assessment; Develop and evaluate quizzes and tests using Moodle modules*. Birmingham, UK: Packt Publishing.
- [5] Weiss, D. J. (1985). Adaptive Testing by Computer. *Journal of Consulting and Clinical*, 774-789

DEVELOPING AN ACCREDITATION MODEL OF SECONDARY SCHOOL

¹⁾ Marjuki, ²⁾ Djemari Mardapi, ³⁾ Badrun Kartowagiran

¹⁾ LPMP of Bengkulu, ^{2),3)} Yogyakarta State University

¹⁾ marjuki.pep@gmail.com, ²⁾ djemarimardapi@gmail.com ³⁾ badrunkw@yahoo.com

Abstract - This study aims to: (1) obtain the accreditation model of secondary school that is effective, efficient, and accountable, and (2) assess the effectiveness on the accreditation model of secondary school. The method of research uses research and development. There are two main activities, namely the development of models and test models. Development of a model aiming to obtain product in the form of accreditation instruments and accreditation procedures is conducted through focus group discussion and expert judgment. The trial models are used to determine the effectiveness of the instrument and the accreditation procedure. The respondents are the principal, teachers, and administrative staff of the school. The data on expert assessment results is analyzed by calculating the content validity index using a scale of Aiken's V, while the reliability of estimates is analyzed by intraclass correlation coefficient (ICC). The data on respondents' assessment results is analyzed by calculating a score of effectiveness. The results show that the product accreditation instruments have content validity of 0.81 in a valid category and reliability of 0.75 in a reliable category. The average score of the effectiveness on the instruments of accreditation is 3.40, which is in a good category. The average score of the effectiveness on the accreditation procedures is 3.38, which is in a good category. Based on the assessment of respondents can be concluded that the product of the accreditation model of secondary school is called as The Accreditation Model of Secondary School Real and Trusted abbreviated with The Accreditation Model of SSRT. The Accreditation Model of SSRT is effective, efficient, and accountable that can be used to assess and evaluate the quality of secondary school.

Keywords: *accreditation model, accreditation instruments, accreditation procedures*

I. INTRODUCTION

The benchmark for the quality of education at primary and secondary levels of education is the national standard of education (NSE). According to Law No. 20 of 2003, the NSE is the minimum criteria regarding the educational system in the entire territory of the Unitary Republic of Indonesia. NSE serves as a basis for planning, implementation, and monitoring of education in order to realize the national education quality.

In order to improve the quality of national education, the government has made various efforts such as the development and improvement of curriculum, development of learning materials, improving the evaluation system, procurement of books funding instruments lessons, improvement of infrastructure of education, improving teacher competence, as well as improving the quality of school leadership. However, these efforts until now have not shown results, as expected.

The focus on efforts improving the quality of education is to create an atmosphere of learning and the learning process so that learners can actively develop her potential. Mardapi (2003, p.8) says that improving the quality of education could be achieved through improving the quality of learning and quality assessment system. Increasing the quality of teaching carried out at various levels of education will be able to improve the quality of education.

Assurance and quality control of education in order to improve the quality of education can be done through accreditation. School accreditation is basically an evaluation which aims to

determine the performance, feasibility, or the quality of schools with reference to national education standards. The object of evaluation in the school accreditation process can be used to describe the characteristics of the quality of schools, among others through the information about the physical condition of the school, the curriculum used, school personnel, socio-economic background of students, and the community's role.

The characteristics of the model of accreditation according Sumarno (2000, pp.4-5) as follows: (1) the evaluation process focuses on quality processes and outcomes, not necessarily covers all the detailed information about the school; (2) using a simple scoring system, but presenting things that are substantive and not just administrative; (3) the assessment including key indicators are measurable and observable; and (4) using an instrument that allows the schools do a self-evaluation (self-evaluation) resulting in a process of empowerment schools for purposes of quality improvement on an ongoing basis.

Identification of problems related to the implementation of the model for accrediting schools at the secondary school level, among others: (1) the accreditation instruments have not been used for the assessment quality of secondary school in an integrated manner, (2) items of statement in the accreditation instruments for the components of competency, content, process, and assessment standards have not been developed in accordance with the latest regulations, (3) evidence of the compliance component of national education standards is more dominant in the form of administrative documents, (4) the majority of stakeholders considers that the accreditation results are less accurate and less credible, and (5) the majority of school has difficulties in filling the instruments of collecting data and supporting information.

Results for accrediting schools other than specified by the quality of the instruments and procedures used also depends on how the quality of assessors in the assessment. Accreditation and assessment are two very important process in maintaining the quality of education units (Patil and Pudlowski, 2005, p.49). In the accreditation process usually begins with the assessment. The assessment results are then used as a basis for giving legal form of accreditation status.

Pagliarulo (1996, pp.1114-1115) argued that the accreditation is the process whereby an organization recognize the institution include the skills possessed by certain standards. The accreditation meaning is almost the same with the understanding accreditation imposed in schools. There are three aspects of accreditation: (1) the accreditation process conducted by an external organization or institution, (2) units which can be accredited program of study or educational unit, and (3) certain standards must be determined first.

In the Accreditation Handbook, COA (2008, p.3) accreditation formulated as follows:

Accreditation is defined as verifying that the sponsorship program in question has demonstrated that, when judged as a whole, it meets or exceeds the Commission's Common and Program Standards Adopted as selected by the program sponsor pursuant to the options listed in the Accreditation Framework. The program sponsor (including its credential programs) is judged to be effective in preparing educators and is demonstrating overall quality in its programs and general operations. The status of Accreditation can be Achieved even if there are one or two Common standards identified as met with concerns or one or more areas of concern are identified within credential programs.

Accreditation in the context of the quality assurance on education can be interpreted as an external evaluation. According to ECA (2005, p.4), accreditation is seen as an independent and a formal decision obtained from an independent body based on certain standards. ECA stated:

Accreditation is achieved through a multi-step process (selfevaluation/documentation submitted by the unit undergoing accreditation; external assessment by independent experts; the accreditation decision). The accreditation decision depends upon a quality assessment based on internationally accepted quality standards. The final decision of the accreditation procedure itself is authoritative in nature, has been determined by an external process, and results in a "yes" or "no" judgment with a limited validity.

In the process of accreditation of a school with activity measurement, assessment, and evaluation. According to Huitt (1996, p.1), the measurement is the quantification of data, assessment is the process of collecting data to understand an issue or object better, and the evaluation is the process of comparing the data available with standard or criteria established. While Stanley and Hopkins (1978, p.3) argued that the evaluation is used in a broader context

and can be carried out both externally and internally, while the assessment is used in the context of a more narrow and generally implemented internally.

Accreditation is an evaluation process that is essential in order to maintain the quality of education (Patil, 2005, pp.49-58). There are three stages of the educational process, namely: input, process, and output. Accreditation is essentially a form of evaluation of the evaluation of the performance standards or criteria, the quality of education. The purpose of evaluation is to see whether the program that has been planned is reached or not, valuable or not, and efficient or not in practice. Evaluation is associated with the value judgment.

In a school with accreditation assessment and evaluation activities. Assessment is done based on the evidence in accordance with the criteria or accreditation standards. In line with this, Mardapi (2008, p.1) stated that the assessment includes the collection of evidence about the achievement of a program related to all of the educational process, such as: curriculum, facilities and infrastructure, administration, teaching methods, and learners, whereas evaluation is one of a series of activities to improve the quality, performance or productivity of an institution in implementing the program.

Recommendations on the accreditation results of the school can be used for the basis of decision making in the effort to improve school quality. In this context, there is a school in the accreditation evaluation activities. In line with this, Kartowagiran (2006, p.1) said that the evaluation is a method to identify and assess the effectiveness of a program by comparing the predetermined criteria or objectives to be achieved with the results achieved. The results achieved in the form of information, used as consideration for decision-making and policy-making.

School accreditation can be viewed as a quality evaluation of school. Evaluation model commonly applied for the purposes of school accreditation is an evaluation model professional judgment. The basic assumption used is "Best evaluation is the expert opinion of a qualified professional" (Arcraft, 2005, p.133).

The best evaluations are expert opinions of qualified professionals. Evaluation of personal model of this judgment focuses attention on the educational process with the use of professional judgment and the development of standards for the education program. Evaluation model can be called a model of personal judgment SAVE namely: Self-study, Visitation, Annual Report, and Evaluation panel.

School quality evaluation must be done in a comprehensive manner involving all aspects in efforts to achieve educational goals that have been set. Thorndike, Cunningham, and Hagen (1991, p.58) stated that the purpose and usefulness of evaluation in education directed to decisions regarding: (a) teaching, (b) the results of learning, (c) the diagnosis and remedial efforts, (d) placement, (e) selection, (f) guidance and counseling, (g) curriculum, and (h) institutional assessment.

School accreditation is one form of quality assurance. Accreditation results of school can describe the quality of the school or the quality of educational programs. Lindsay and Campbell (2002, pp.29-31) tested the accreditation status as one indicator of the quality of the education program stating that the accreditation status is often regarded as an indicator of the quality of programs. Bennett (2002, pp.1-6) concluded that the school's accreditation status can be used as guidelines for selecting good quality of schools. Accreditation may initially raise the general quality of assurance, with a specific bit of information about the school.

Based on the above, the researchers have found that school accreditation is very important to know the quality of secondary school. In order to obtain information about the quality and comprehensive secondary school that is valid, then the existing accreditation models need to be developed through research in order to obtain accreditation model of secondary school better. In connection with that, the problem in this research is: How to obtain an accreditation model of secondary school that is effective, efficient, and accountable?

In harmony with the formulation of the problem, then the purpose of this research is to obtain an accreditation model that is effective, efficient and accountable. There are two benefits of this research, which can be used as a reference or benchmark in the development of the theory or model of evaluation of education and can be used to assess or evaluate the quality of secondary school either through accreditation, quality auditor or quality mapping.

II. RESEARCH METHODS

The method used in this study is a modification of the combined model of research, development and diffusion of Hopkins and Clark and models of research and development of Borg & Gall. Research and development phases of this research are: (1) the study of the problems and data collection, (2) development planning, (3) development of draft models, (4) the trial phase 1 and the revised phase 1, (5) the trial phase 2 and revision phase 2, (6) the test on the effectiveness of the model, and (7) the preparation of the final product.

Model of product development is conducted through focus group discussion (FGD). Results on FGD are obtained consensus on component assessment and accreditation standards as a measure of the quality of secondary school. Referring to the accreditation standards that are developed as assessment indicators, assessment techniques, and evidence must be met. It is also developing accreditation, application instruments on the field data of accreditation, technical manual charging instrument of accreditation guidelines for assessment and accreditation procedures. In the second stage, FGD activities have conducted a review on: (1) the formulation of assessment indicators, assessment techniques, and evidence that must be met; (2) a draft accreditation instruments equipped with instruments containing data on the application of the accreditation, technical manual of charging instruments of accreditation and assessment guidelines; and (3) a draft of the accreditation procedures.

The accreditation instruments of product are validated by three experts consisting of experts school accreditation, expert evaluation of education and education management experts. Data expert assessment results are analyzed by calculating the index Aiken's V to determine the validity of the content of the instrument, while the estimated reliability of the instrument was analyzed using SPSS by determining interclass correlation coefficient (Kartowagiran, 2014, p.3).

The accreditation instruments of testing product and the accreditation procedures are done in two stages. Phase 1 trials are conducted in three secondary schools, while the phase 2 trials are conducted in six secondary schools. The target test of secondary school is asked to fill the data and provide information that is relevant to the accreditation instruments. Results of charging data and further information are verified and validated by researchers. In this case, researchers act as the assessors of accredited schools. After filling the instrument of accreditation, the respondents are asked to provide an assessment of the effectiveness of the accreditation instruments that are tested. The respondents are also asked to provide an assessment of the effectiveness of the accreditation procedures.

The techniques of data analyses on the assessments of respondents are done by calculating the average score of the results and assessing the effectiveness of the model. The criteria on the effectiveness of model as follows.

Table 1. Criteria on The Effectiveness of Model

Scores	Category	Conclusions
> 3,4	Very good	It can be emulated
2,8 – 3,4	Good	It can be used
2,2 – 2,8	Enough	It can be used with some revision
1,6 – 2,2	Less good	It can be used after being revised
< 1,6	Very bad	It can't be used

III. RESULTS AND DISCUSSION

A. Accreditation Instruments of Secondary School

The components of quality assessment on secondary school consist of 8 accreditation standards, namely: (1) the standard of competency, (2) the standard of assessment, (3) the standard of learning process, (4) the standard of curriculum content, (5) the standard of teachers and education personnel, (6) the standard of facilities and infrastructure, (7) the standard of management, and (8) the standard of finance. The accreditation instruments are developed based on key indicators, valuation techniques, and physical evidence of assessment.

The accreditation instruments are validated by three experts. The data from the expert judgment then is analyzed by calculating the scores given by each expert. Based on the data from these scores can be calculated an index or coefficient Aiken's V. As it was quoted from Aiken that the index Aiken's V coefficient > 0.50 indicates that the validity of the content items is good. The results showed that the average index of Aiken's V coefficient of 0.81 indicates the content validity of the accreditation instruments is good.

Furthermore, by using an application program SPSS is obtained results of the Anova analyses from the rater 3, while the reliability of the assessment results can be seen from the consistency between the rater (interrater reliability) in its assessment or can be seen from the results of interclass correlation coefficient (ICC) as shown in the following table.

Table 2. Results of ANOVA analysis of the Third Rater

		Sum of Squares	Df	Mean Square	F	Sig
Between People		36.036	54	.667		
Within People	Between Items	.958	2	.479	2.920	.058
	Residual	17.709	108	.164		
	Total	18.667	110	.170		
Total		54.703	164	.334		
Grand Mean = 4.3758						

Table 2 above shows that there is no significant difference ($p = 0.058$) between the average value given by rater 1, rater 2, and rater 3. Thus we can say that the statement items developed by the basis of key indicators, techniques assessment, and physical evidence are valid votes. The estimation of inter-rater reliability can be seen from the results of interclass correlation coefficient in the table following pages.

Table 3. The Results of Interclass Correlation Coefficient (ICC)

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.506 ^a	.348	.651	4.070	54	108	.000
Average Measures	.754 ^c	.616	.849	4.070	54	108	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Table 3 above shows that the inter-rater reliability or interclass correlation coefficient of 0.506 is in a good category and it can be said that the items in the statement of accreditation instruments are in accordance with the key indicators, valuation techniques, and physical evidence assessment.

Research findings show that there is a difference between the accreditation instruments of secondary school with models of The National Accreditation Board for Schools (BAN-S/M) and the accreditation instruments of secondary school with development results. The product characteristics for the accreditation instruments of secondary school and instrument applications contain data on accreditation, (2) the school must complete the instrument application data fields of accreditation before filling accreditation instruments, (3) the proportion between the physical evidence and the administrative evidence shows the factual performance.

B. Accreditation Procedures

Accreditation procedures consist of four stages, namely: (1) the registration accreditation, (2) notification of the visitation, (3) the implementation of the visitation, (4) the submission of reports on the implementation of visitation, (5) validation report of visitation, and (5) the determination of the accreditation results.

Some of the things that are different from the previous accreditation procedures as follows. Registration accreditation done on-line at registration time range has been determined through a portal or web-official accreditation institution. After registering, the school obtains registered receipts for accreditation sent via school email address.

Before the visitation, the accrediting agencies have to validate the implementation of school self-evaluation during the last 3 years. If the analysis shows an increasing trend or consistency on the fulfillment of the accreditation on standard components, the accreditation agencies state that school deserves to be accredited.

C. Trial Results

The results charging the accreditation instruments of product development by the respondents comprise of principals, teachers and education personnel and after the verification and validation by the researchers showing that the value per accreditation standards and accreditation of the final value when compared to the data value of accreditation version The National Accreditation Board for Schools (BAN-S/M) show the same tendency mainstream. According to the product assessment using the results of accreditation instruments. Thus, the product model of secondary school developed by accreditation instruments can be used to assess the quality of secondary school.

The assessment results of product effectiveness on accreditation instruments and the accreditation procedures of secondary school conducted by the respondents obtain the data with the average score of effectiveness as shown in the following table.

Table 4. Average Score of Effectiveness On Accreditation Instruments of Secondary School

No.	Aspects Assessed	Average Score	Category
1	Instructions on filling in model instruments	3,40	Very good
2	Display/design model of the instrument	3,50	Very good
3	The accuracy of the substantive content on data entry	3,50	Very good
4	The validity of the data (data accuracy) of the charged	3,50	Very good
5	Objectivity of data entry	3,40	Good
6	Practicality during data entry	3,30	Good
7	The language used	3,50	Good
8	Feasibility (it may be applied)	3,10	Good
9	Clarity of technical manual	3,40	Good
	Overall of Average Score	3,40	Good

Table 4 shows that the average score of the effectiveness of the instruments accreditation of secondary school in research results as a whole of 3.40 is in a good category including both categories so that the product can be used as the accreditation instruments of secondary school.

Furthermore, the effectiveness of data with the average score of the accreditation procedures of secondary school based on appraisals to the respondents is as shown in the following table.

Table 5. Average Score of Effectiveness on Accreditation Procedures of Secondary School

No.	Aspects Assessed	Average Score	Category
1	Display/design model of procedure	3,60	Very good
2	The accuracy of the substance of the procedure	3,40	Good
3	The language used	3,60	Very good
4	Feasibility (it may be applied)	3,30	Good
5	Procedure accountability	3,00	Good
	Overall of Average Score	3,38	Good

Table 5 shows the average score on the effectiveness of the overall accreditation procedures of secondary school in research results is 3.38 which is in such a good category including both categories that the product can be used as the accreditation procedures of secondary school.

IV. CONCLUSIONS AND SUGGESTIONS

The Accreditation Model of SSRT is effective, efficient, and accountable that can be used to assess and evaluate the quality of secondary school. The accreditation instruments of secondary school have content validity index based on Aiken's V of 0.81 in a valid category and reliability by interclass correlation coefficient (ICC) of 0.75 in a reliable category. The average score of the effectiveness on the instruments of accreditation is 3.40, which is in a good category. The average score of the effectiveness on the accreditation procedures is 3.38, which is in a good category.

Suggestions to: (1) the head of department of education in province/district can socialize the Accreditation Model of SSRT for supervisors, principals, teachers, administrators, and school committee, (2) the head of the board on quality assurance of education (LPMP) is to be able to socialize The Accreditation Model of SSRT and facilitate the training on the use of accreditation instruments to the accreditation assessors, supervisors, principals and teachers for the purpose of mapping the quality of secondary school and school self-evaluation, and (3) the head of

secondary school and the school's development team can use the accreditation instruments to conduct school self-evaluation and make preparation to be accredited.

BIBLIOGRAPHY

- [1] Artcraft. (2005). *Program evaluation*. Sydney: Artcraft Research.
- [2] Bennett, W.M. (2000). *Independent school accreditation: A guide to school choice*. Accessed on October 22, 2008, from <http://www.educationalconsulting.org/PDF/Independent/School/Accreditation.pdf.htm>
- [3] Huitt, W. (1996). *Assessment, measurement, and evaluation: Undergraduate version, educational psychology interactive*. Diambil pada tanggal 20 Juni 2009, dari <http://chiron.valdosta.edu/whuitt/edpsyc/edpmsevl.html>
- [4] Kartowagiran, B. (2006). *The basic principles of monitoring and implementation*. The training material of Team Implementation Monitoring and Evaluation of DKI Jakarta Province. Yogyakarta: Yogyakarta State University.
- [5] Kartowagiran, B. (2014). *Optimization of the level of competence in vocational testing to improve the soft skills of graduates*. The seminar papers. Yogyakarta: Yogyakarta State University.
- [6] Lindsay, D.H. & Campbell, A. (2002). *An examination of AACSB accreditation status an accounting program quality indikator*. Accessed on October 22, 2008, from <http://www.sbaer.uca.edu/research/TemporarilyDisabled-wdsi/2002/pdffiles/papers/012.pdf.htm>
- [7] Mardapi, D. (2003). *Design and assessment of student learning*. The paper presented at the workshop on quality assurance system of learning process, dated June 19, 2003 at the Gadjah Mada University.
- [8] Mardapi, D. (2008). *Technique making instrumens: test and non-test*. Yogyakarta: Mitra Cendekia Press.
- [9] Pagliarulo. (1996). Accreditation: Its Nature, Process, and Effective Implementation. *Physical Therapy Education Journal*, 66 (7), 1114-1118.
- [10] Patil, A.S. & Pudlowski, Z.J. (2005). Important issues of the accreditation and quality assurance and a strategy in the development of an accreditation framework for engineering Courses. *Global Journal of Engineering Education*, 9, 49-58.
- [11] Stanley, J.C. & Hopkins, K.D. (1978). *Educational and psychological measurement and evaluation*. New Delhi: Indiana Prentice-Hall of India Private Limited.
- [12] Stufflebeam, D.L. & Shinkfield, A.J. (2007). *Evaluation, Theory, Models, and Applications*. San Francisco: John Wiley & Sons, Inc.
- [13] Thorndike, R.M., Cuningham, G.K., Thorndike, R.L., Hagen, E.P. (1991). *Measurement and evaluation in psychology and education*. New York: Macmillan Publishing Company.

Developing an Instrument for Assessing the Performance of High School Physics Teacher

Nurul Fitriyah Sulaeman¹, Badrun Kartowagiran²

1. Departement of Physics Education, Mulawarman University

2. Graduate School, Yogyakarta State University
nurulsulaeman@yahoo.com

Abstract- The aims this research to reveal: (1) the constructs of instruments, (2) the characteristics of the instrument seen by colleagues, (3) the characteristics of the instrument seen by student, (4) the characteristic similarity between the instrument seen by colleagues and the instrument seen by students. This research and development which develops an instrument for assessing high school physics teachers' performance by collagues and students. The streps of instrument development are: arranging the instrument spesification, writing the instrument, determining the instrument scale, determining the scoring and norm system, reanalysing the instrument by experts, try out the instrument, analysing the inatrument, organizing the instrument, corelating the score of colleague instrument and student questionnaire. The subject of the trial were 30 high school physics teacher assessed by a colleague using observation sheet and 450 students using a questionnaire. The instrument validation was estimated through expert judgement and factor analysis, while the reliability estimation used Cronbach Alpha coefficient. The result shows that: (1) the components that form the instrument include lab work learning plan, lab work learning preparation, lab work implementation, tools administration and lab work result; (2) the validation of the instrument bu colleagues using expert judgement and factor analysis rated good with the loading factor is 0.432 – 0.900 and the realibility is 0.887 rated high; (3) the validation of the instrument by students using expert judgement and factor analysis rated good with the loading factor is 0.384 – 0.746 and the realibility is 0,935 rated high; (4) observation sheet seen by colleagues and students questionnaire have same characteristics.

Keywords: *development, instrument, performance, physics teachers*

I. INTRODUCTION

Policymakers, parents and practitioners agree that the key to improving public education in Indonesia is placing highly skilled and effective teachers in all classrooms. Yet the nation still lacks a practical set of standards and assessments that can guarantee that teachers are well prepared and ready to teach to build a better education. According to Clotfelter (Darling-Hammond, 2010) Teachers can have large effects on student achievement, as suggested by a recent large-scale study in North Carolina, which found that the differences in achievement gains for students who had the most qualified teachers versus those who had the least qualified were greater than the influences of race and parent education combined. The results showed that teachers who have low skill can affect the low success of student learning.

Increased attention for teachers from the government continued with increase in the demands on the ability of a teacher should possess. In performing its duties, the teacher must have the expertise, specialized skills that meet certain quality standards or norms are acquired through professional education. Teachers as a professional job should continue to improve its ability to manage teaching and learning activities.

According to Resnick & Resnick (Johnson et al, 2009: 18), the assessment can be built in accordance with what is to be taught, because we will get what we value. Before the certification system exists, the teachers are not accustomed rated by performance. So that the good performance of the teachers will be difficult to measure because there are no standardized assessment. The performance appraisal directly assess the ability of teachers when teaching can provide another illustration of the ability of teachers that can be input for further improvement.

The instruments that suitable for asses bsing the performance of teachers of physics lab is also not yet available, it is necessary to develop assessment instruments physics lab teacher

performance has been tested empirically and easy to use. This paper will reveal the construction of assessment instruments lab school physics teacher performance by peers and students then find out the similarities between the characteristics of both instruments.

The task of teachers in teaching practice activities can be observed in Figure 1.

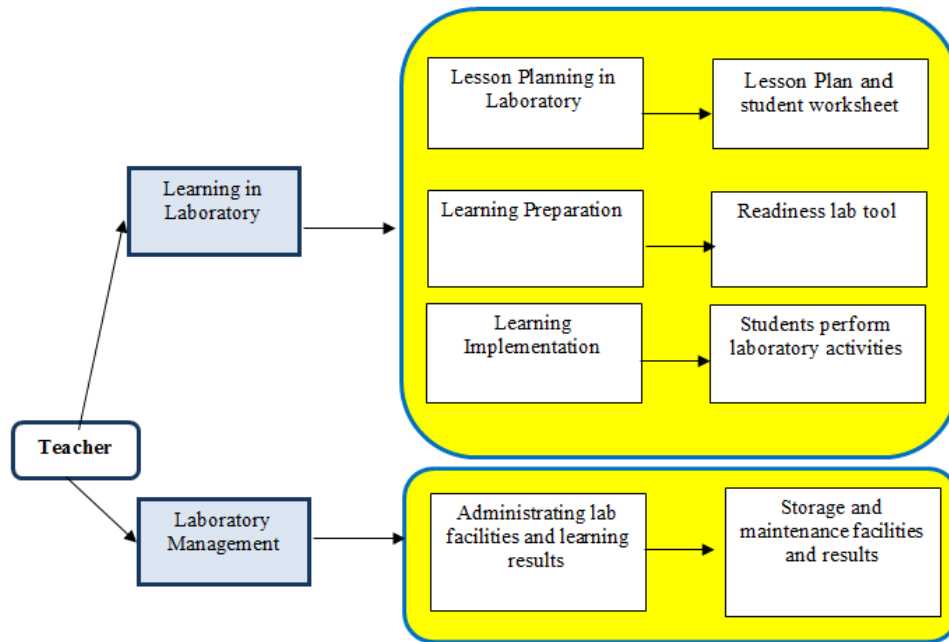


Figure 1.

II. RESEARCH METHODOLOGY

This study is a research & development which develop assessment instruments lab school physics teacher. Model development of assessment instruments this school physics lab follow the development of non-test instrument according to Djemari Mardapi (2008: 108) is modified.

Specifications instruments developed in this study is an assessment instrument lab physics teacher performance. Teacher performance assessment instruments developed consists of two instruments, first instrument rated by collague and the second rated students. Instruments blue print can be observed in Table 1.

Table 1. Dimensions and Practical Physics Teacher Performance Indicators

Number	Dimension	Indicators
1.	Lesson Planning in Laboratory	Learning topics and the formulation of objectives
		Work analysis
		Outline learning materials determination
		Lab activity plan
		Outline learning
2.	Learning laboratory preparation	Equipment preparation
		Lab materials preparation
		Lab support preparation
3.	Learning Implementation	Opening of learning
		Initial Lab activity
		Lab activity
		Closing of learning
		Evaluation
4.	Laboratory Management	Tools administration
		Materials administration
		Lab document administration
		Lab results administration

Instrument lab school physics teacher assessment development procedures can be observed in Figure 2.

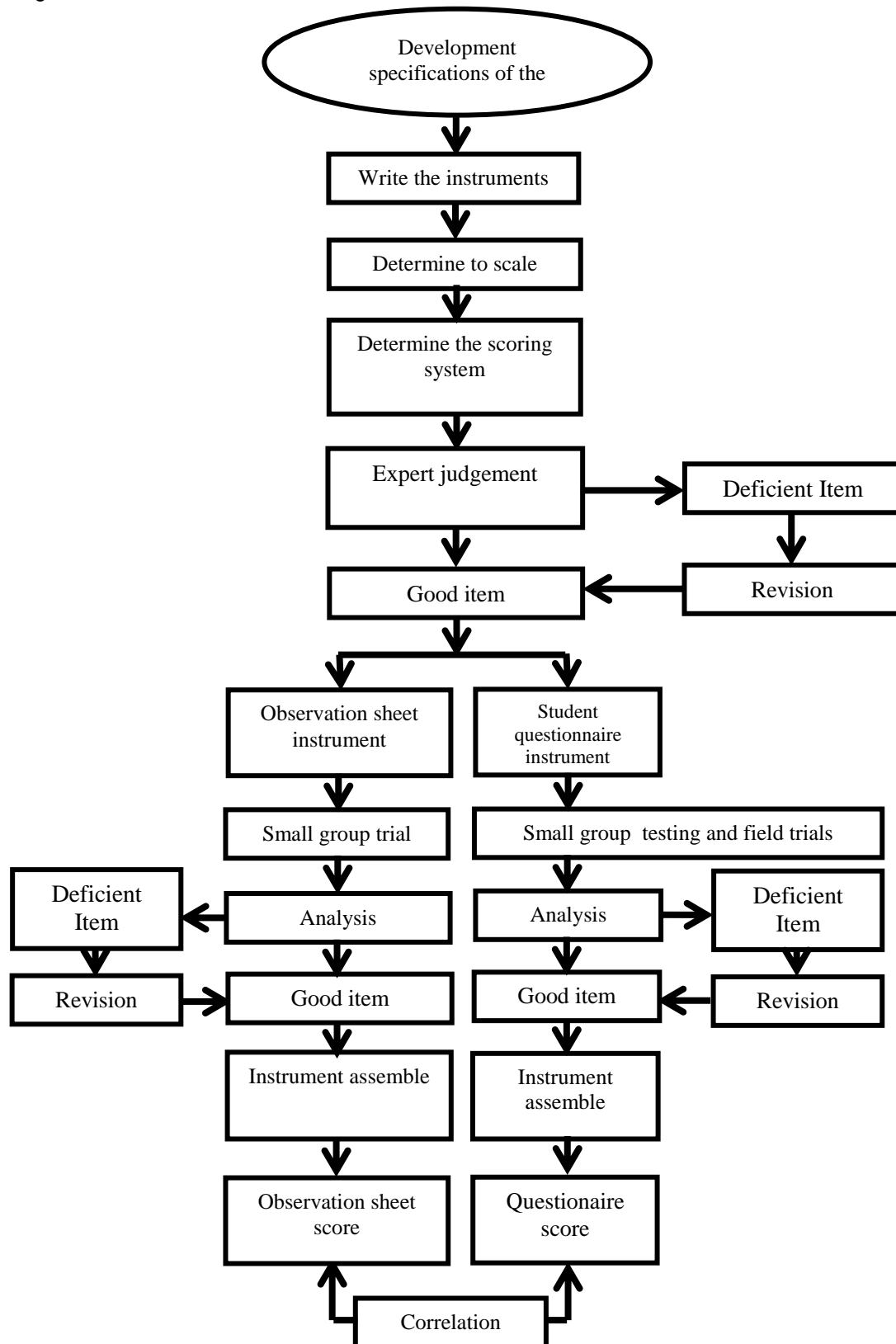


Figure 2. Research procedure

III. RESULTS AND DISCUSSION

A. Expert Judgement

The initial draft of this instrument in the form of blueprint physics lab teacher performance assessment which will be developed into a major instrument (observation sheet by colleague) and replacement instruments (questionnaires perception by students) which have the same items. Expert judgment phase is an initial phase to validate the preliminary draft that has been developed. This stage involves five experts in validating these instruments, experts A as an expert in the field of educational measurement, expert B and C as an expert in the field of physics learning, and expert D and E as the measurement expert in the field of physics learning. Planning aspects of physics lab activities are not included in the student assessment instruments with the consideration that the students do not know how to plan a physics Lab activity.

Selection of observations consists of four options A, B, C, and D should be added to the answer option E which is open choices where colleagues and students can write the observation situation that different with the option. This is done to anticipate if at the time of observation there is no option in accordance with the activities undertaken by teachers.

B. Trial Phase

Trial of the instrument done in two stages in high school in the province of D.I. Yogyakarta which conducted in the second semester of the 2011/2012 academic year between March to June 2012. Small group trial conducted in SMA Negeri 1 Sedayu Bantul on 23 April 2012 with 10 people attended the tenth grade students who have attended the Physics lab activities. There is some feedback from this activities to improve our instruments developed.

The field trials involving 450 high school students of class X and XI in the province of Yogyakarta D.I who has followed the physics lab activities and 30 physics teachers who do the learning lab. Respondents were selected from 25 secondary schools of public and private. Each regencies is represented by 5 schools consisting of three public schools and two private schools.

C. Data analysis

The results of calculations can be observed in Table 2 which shows that the main instrument and the replacement has a high reliability or have a very good level of reliability to be used as an assessment instrument lab school physics teacher performance.

Table 2. Estimation of Reliability

	Instrumen	Cronbach's Alpha	N of Items
Main	Total	0,887	42
	Planning assessment	0,802	15
	Observation sheet	0,945	27
Replacement	Questionnaire perception	0,935	27

Empirical validity test performed using factor analysis. Factor analysis aims to perform data reduction into several key factors. In this study, the main and replacement instruments consists of 27 items together that analyzed using factor analysis. The initial stage of the analysis is to test variable factors which are considered eligible to be included in subsequent analyzes. This testing is done with all the items together in the main instruments and replacement instruments. The items have a tendency to cluster and form a factor then the variable will have a fairly high correlation with other item.

The test results for the KMO and Bartlett's main instrument KMO-MSA value of $0.441 < 0.5$ so that the value of KMO-MSA can not be accepted. Bartlett's test of significance value $0.000 < 0.05$, which indicates that these variables are correlated. The next step needs to be observed on the table Anti-image Matrices. Item B7 has a coefficient of 0.120 anti image that shows low correlation. This causes the B7 item removed or not included in further analysis. Others items remain excluded from further testing as anti image has a coefficient of > 0.3 . After item B7 issued, performed the analysis again. The test results re-KMO and Bartlett's show the value of KMO-MSA amounted to $0.649 \geq 0.5$ includes both categories so that it can be further analyzed using factor analysis. The test results KMO and Bartlett's in replacement instruments can be observed in Table 8. KMO-MSA value amounted to $0.942 \geq 0.5$ so that variables can still be predicted

without error and can be further analyzed using factor analysis. Bartlett's test of significance value $0.000 < 0.05$, which indicates that these variables are correlated. Both these conditions indicate that the factor analysis can be continued.

Number of factors formed can be determined by looking at the output of factor analysis on the part of the total variance explained table but still have to consider the factors specified in the early development instrument. Main instruments factors can be observed in Table 3 which shows that the three main factors that are formed. It can be concluded from the 26 items extracted into three dominant factor. These three components are able to explain the total variance of 64.06%.

Table 3. Total Varian That Explained by Main Instrumen

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	11,261	43,310	43,310
2	4,042	15,545	58,856
3	1,637	6,296	65,152
Extraction Method: Principal Component Analysis.			

Results faktoring main instrument supported by the results factor replacement instrument that can be observed in Table 4. Component 1 is the dominant factor because it has the largest eigenvalue is 10.543. Component 2 has eigenvalue 1,694 and components 3 at 1,487. It can be concluded from the 27 items extracted into three dominant components. These three components are able to explain the total variance of 50.83%.

Table 4. Total Varian That Explained by Replacement Instrumen

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	10,543	39,048	39,048
2	1,694	6,275	45,323
3	1,487	5,508	50,832
Extraction Method: Principal Component Analysis.			

Determination of the number of factors that are formed from grains of instruments made by considering eigenvalue, drawings screeplot and construct the initial underlying instrument development. Validation of the content that has been done using expert judgment to produce the same 3 factors in shaping the main instrument and replacement. The next step after a number of unknown factors is the determining factor pair with the appropriate grain. To facilitate interpretation of the components that form the rotation performed using varimax rotation.

The results of the factor analysis performed on the main and replacement instruments shows B7 invalid item so that the item is discarded. In the main instruments of all item are valid and appropriate measure factors to be measured with the loading factor in the range between 0,438 to 0,900. On the replacement instruments, there are two items that shifting factors need to be revised so that the grain B6 and D3 while other items are in accordance factor diinginkan. Loading factor for grain replacement instruments ranging from 0,384 to 0,746. Replacement instrument is a perception questionnaire students so there is a possibility there are grains of elusive or ambiguous for students. Once revised, these points remain stationed on the factors corresponding to the initial construct.

The assessment criteria using the criteria of BSNP stating the performance of teachers considered eligible performance is considered good if the score reaches $\geq 70\%$ of maximum score. Each item has a maximum score of 4 and a minimum score of 0. Sheet assessment learning plan consists of 15 items that have a maximum score of 60. At the observation sheet by peers and students perception questionnaire consisting of 26 items, the maximum score is 104 so that the teacher's performance included either category if more than 72.

To determine the possibility of a perception questionnaire instrument can replace the role of student observation sheet instruments by colleagues estimate the correlation coefficient using the Pearson product moment correlation coefficient. Correlation is done at an average score of student questionnaire with a score ratings by peers, the calculations are complete can be observed in Table 5. These results show a correlation between the scores of instruments colleagues and students is high.

Table 5. Correlation Score Assessment Colleague and Students

Students Score		Colleague Score		
		Preparation	Implementation	Administration
	Preparation	0,709		
	Implementation		0,919	
	Administration			0,732
	Total	0,935		

IV. CONCLUSION

From the development of assessment instruments lab school physics teacher performance can be summarized as follows.

1. Constructs practicum teacher performance assessment instrument consisting of a high school physics lesson planning factor lab, lab lesson preparation, implementation of practical teaching, administrating means and lab results.
2. Validation peer assessment instrument by using expert judgment and factor analysis is quite good with a loading factor ranging between 0,438 to 0,900 and the high reliability that is equal to 0.887.
3. Validation of student assessment instruments by using expert judgment and factor analysis is quite good with a loading factor ranging between 0,384 to 0,746 and the high reliability that is equal to 0.935.
4. Instruments colleague observation sheets and questionnaires perception students have the same characteristics and a high correlation scores so students can replace the perception questionnaire observation sheet peers.

REFERENCE

- [1] Badan Standar Nasional Pendidikan. (Maret 2012). Paradigma Pendidikan Nasional. Jakarta: BSNP
- [2] Chiappeta, E.L. and Koballa, T.R. (2010). *Science Instruction in the middle and secondary schools*. Boston : Allyn & Bacon.
- [3] Darling-Hammond, L. (2010). *Evaluating teacher effectiveness*. http://www.google.co.id/url?sa=t&rct=j&q=teacher%20performance%20assessment&source=web&cd=4&ved=0CD4QFjAD&url=http%3A%2F%2Fwww.americanprogress.org%2Fissues%2F2010%2F10%2Fpdf%2Fteacher_effectiveness.pdf&ei=ZecMT7nhKMHprAfd79i7BA&usg=AFQjCNEjcRWpbZhkIWBMDEHqke0Yhl772Q&cad=rja. Diambil tanggal 11 Januari 2011.
- [4] Departemen Pendidikan Nasional. (2005). Undang-undang Republik Indonesia Nomor 14, Tahun 2005, tentang Guru dan Dosen.
- [5] Departemen Pendidikan Nasional. (2008). Penilaian Kinerja Guru. Jakarta : Direktorat Tenaga Kependidikan.
- [6] Djemari Mardapi. (2008). Teknik penyusunan instrumen tes dan nontes. Yogyakarta : Mitra Cendekia Press.
- [7] Groundlund, N. E., (1976). *Measurement & evaluation in teaching*. New York: Macmillan Publishing Co., Inc.
- [8] Jonathan Crowther. (1995). *Oxford advanced learner's dictionary*. Bristol : Oxford University Press.
- [9] Jonson, R. L., dkk. (2009). *Assessing performance, designing, scoring, and validating performance tasks*. New York: The Guildord Press.
- [10] Miller, P. W., (2010). *Measurement and teaching*. Indiana: Pwmleronline.
- [11] Paul Suparno. (2007). Metodologi pembelajaran fisika. Yogyakarta: Universitas Sanata Dharma.
- [12] Purwanto. (2007). Instrumen penelitian sosial dan pendidikan. Yogyakarta : Pustaka Pelajar.
- [13] Peraturan Menteri Pendidikan no 16. (2007). Standar kualifikasi akademik dan kompetensi guru. Jakarta : Depdiknas.
- [14] Soeprijanto. (2010). Pengukuran kinerja guru praktik kejuruan. Jakarta : CV. Tursina.

ANALYSIS INSTRUMENTS TEST READING FOR ACADEMIC PURPOSE STUDENTS OF ENGLISH EDUCATION UNISNU JEPARA

Nusrotus Sa'idah ¹⁾, Hayu Dian Yulistianti ²⁾

^{1),2)} FTIK UNISNU Jepara

¹⁾ nusrotussaidah17@gmail.com, ²⁾ hayuhasan07@gmail.com

Abstract - The research is quantitative descriptive. The aims of the study to determine the characteristics of the questions test were analyzed quantitatively, reading for students of English language education. The test was focused on WAP as the main competency of reading for EAP and elaborated sub-competencies into literal, inferential, and critical reading. This study is to improve the competence of lecturers in developing the test instrument. Samples from this study is the reading test of final exam, consist of 10 questions in Multiple Choice, 10 in True False and 5 in open- response questions. The source of the data is taken from answer sheets of 91 English students and the answer key. Quantitative analysis used manual Microsoft excel and ITEMAN 3:00 versions. Problem was analyzed according to the validity, the reliability, item of difficulty, and effectiveness of the destructor. The instrument has a relatively low reliability of the test is 0.671. The results showed the characteristics of multiple-choice questions for the difficulty items ranged from 0.215 up to 0.957, discriminating power ranged from 0.291 up to 0.698. 90% of multiple choice questions are accepted. In the True false questions used manual Microsoft Excel, showed in discriminating power ranged from -0.19 to 0.56. In the item of difficulty ranged from 0.131 to 0.868. The instrument true false has a relatively low reliability of the test is 0.252. The false true questions are 70% accepted, viewed in the item of difficulty and discriminating power.

Keywords : test analysis , reading for academic purpose, reliability

I. INTRODUCTION

The quality of education in Indonesia is still low, one of the reasons is the student's ability to improve cognitive, affective and psychomotor domain. The expectation of change in order to improve and increase the quality of learning through assessment paradigm is a must in our education.

Evaluasi merupakan salah satu cara dari beberapa rangkaian kegiatan untuk meningkatkan kualitas, kinerja, atau produktivitas suatu lembaga untuk melaksanakan program (Mardapi, 2008). Evaluation includes two activities, there are measurement and assessment. Measuring is an activity of pairing facts of a certain size. While the assessment is to provide a decision is good or not. In the evaluation activities of learning evaluation to know the competence of the students are already understand the material or not by measuring. The measurement is an activity to assess student's ability in understanding materials and student competence is the achievement test.

Assessment is any systematic procedure for collecting information that can be used to make inferences about characteristics of people or object (AERA et al., 1999). This assessment is very important activity that cannot be separated in education. Good assessment system will be able to give an idea of the quality of learning that is seen from the results of the assessment, while students as a reference in increasing motivation to learn. To judge something through measurements, a form of measurement is testing. Test in the field of education known as a test.

Test is a group of question students choose and answer or assignment that students should do to measure the competence. According to Lee J. Cronbach (Saifudin Azwar, 2010:2) test is a systemic procedure for observing a person's behavior and describing it with the aid of numerical

scale or a category system. Test is used to know students behavior. The teachers should know foundation of test in order to get a good achievement in test. Therefore, the requirement of a measurement is used to learning achievement test.

Validity as the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of the test ((AERA et al., 1999). Second, consistency is accuracy in measurement. In learning achievement tests, questions of a test should not be too far below or above the level of achievement of the ability to learn and level of difficulty items.

English language competency tests for *English for Academic Purpose* (EAP) have not been formulated by English language learning in particular. In this test the lecturer made a test in accordance with competency standards in reading subjects. The reading tests for students of English language program. The quality of the instrument that has made in this faculty need to be reviewed of the absence of test construction that needs to be stated, this test meets the requirements of the quality of the test or not.

One of skill in English that students should learn is reading. Reading test is needed for students, it focuses on the skills and knowledge of education. To help students become interactive readers, the teacher should create an appropriate design when they make the forms of questions. In mastering reading test, the students not only how the memorize facts and definitions well, but also they have to know main idea of the text and to promote an understanding the text. In developing reading questions in foreign- language, well- designed help students to construct the meaning and create when they interact with the text.

According to our curriculum, multiple- choice and open- response test questions is still used to measure reading comprehension (open-responses is students answer the question by their own word). Readers making meaning from multiple choice reading texts use many strategies in attempting to arrive at the 'correct' meaning as identified by item writers. However, the success of a reader taking a multiple choice reading test is not necessarily a measure of the depth of textual understanding (Thissen, Steinberg & Fitzpatrick, 1989; Kubisyn & Borich, 1987; Farr, Pritchard & Smitten, 1990; Kemp, 1985; Ben-Shakbar & Sinai, 1991; Bridgeman, 1992; Seda, 1989; Bauman, 1982 in Naughton 1996). True and false question is one part of the structure in the reading test. In doing this test, students should understand sentences and identify whether the statement of the sentence is appropriate to the text or not. If the statements are appropriate with the text, the student answers the question *true*, if it does not match then the student answers *false*.

II. RESEARCH METHODOLOGY

This research used descriptive exploratory studies, for exploring the characteristics of a reading test in English Academic Purpose (EAP) quantitatively. Analysis of quantitative data's in manual calculations for instrumental analysis in the form of true- false. Calculation in multiple choice questions of instruments using classical test theory. It is used to English students test in UNISNU Jepara.

The research data is taken from the answer sheets of the results of English student's exam that followed the reading course along with the test device. The sample was 91 students from three group classes.

To collected this data used documentation from syllabus, latticework, the test device and the answers of the students after the examination. Steps in analyzing the reading test questions are as follows

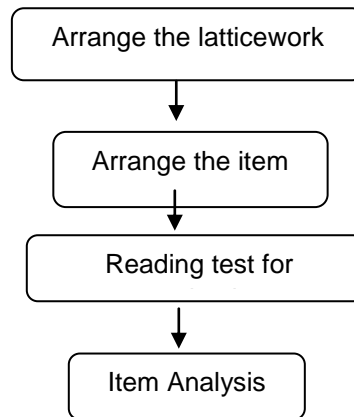


Figure 1. Procedure Analysis item reading test

Quantitative data analysis with classical test theory assisted by ITEMAN 3.00 version and Microsoft Excel. Some aspects were analyzed quantitatively. There are reliability, item I of difficulty, discriminating power and the effectiveness of destructor.

A. Reliability

The instrument can be reliable if the test will provide consistent information on the same subject although at different times. The instrument test used in this study is multiple choices, true false and dichotomous essay. In this study, the reliability values obtained from the analysis ITEMAN 3.0. Reliability value can be calculated using the formula KR 20:

$$r_{11} = \left(\frac{R}{R-1} \right) \left(\frac{\sigma_i^2 - \sum pq}{\sigma_x^2} \right) \dots \dots \dots (1)$$

R is the number of items, σ_i^2 is the kinds of item i and σ_x^2 is the kinds of total score. It means that the higher score, the greater total of coefficient in reliability, and the greater the degree of accuracy of the test in measuring the ability of learners.

B. Item Difficulty Level

Item difficulty level is defined as the percentage or proportion of test takers who correctly answer item (Reynolds,). The item of difficulty depends on the purpose of the test. The test is intended to compare someone capable of high, medium or low. Difficulty index of the simplest item and most widely used is the difficulty level of P, which is the number of correctly answering the item to number of examines.

The equation used to determine the level of difficulty (p):

$$P = \frac{\sum B}{k} \dots \dots \dots (2)$$

P : the proportion of items answered correctly on certain
 $\sum B$: the number of test participants answered correctly
 N : the number of test takers

Menurut Djemari Mardapi (2008:116) besarnya tingkat kesukaran yang dapat diterima adalah 0,30 sampai 0,80. Index difficulty is 0.3 showed hard item, however index difficulty is 0.8 showed item easy. Item difficulty showed high is item aesy.

C. Discriminating Power

Daya beda soal diukur dari kesesuaian soal tersebut dengan keseluruhan tes dalam membedakan antara peserta tes yang berkemampuan tinggi dan yang berkemampuan rendah dalam hal yang diukur oleh tes yang bersangkutan (Sumadi Suryabrata, 2005:131). The techniques used in measuring the power of difference is the correlation between scores on a particular matter (the correlation formula biserial)

$$r_{pb} = \frac{M_p - M_t}{s_t} \sqrt{\frac{p}{q}} \dots\dots\dots(3)$$

- r_{pb} : the correlation coefficient point biserial
 M_p : the mean score well on the test answers
 M_t : the average total score (all participants test)
 p : the proportion dichotomy 1 (correct answer)
 q : the proportion dichotomy 0 (wrong answer)
 SD : the standard deviation of total score

According to Sumadi Suryabrata (2005:131) *butir yang diterima, indeks daya bedanya $\geq 0,3$. Butir indeks daya beda lebih besar dari 0,1 dan lebih kecil dari 0,3 perlu direvisi. Butir ditolak bila indeks daya bedanya kurang dari 0,1. If the item discriminating between students who know the material and those who do not, then students who are more knowledgeable. Should get the item correct more often than students who are less knowledgeable.*

D. The effectiveness of Destructor

The destructor is good if it has negative value point biserial correlation. A negative value indicates that the low-ability students tend to choose that destructor as an answer and instead high-ability students will select an answer key as the answer. According Djemari (2008: 143) said the destructors functions have at least 3% of students.

III. RESULT AND DISCUSSIONS

Multiple choices in reading test used classical test theory of the program ITEMAN 3:00. The results of test analysis include: reliability, item of difficulty, discriminating power and effectiveness of each item on the destructor. From the analysis of multiple choices with 3:00 ITEMAN, as follows:

Table 1. Statistic scale of reading test

Number of items	10
Number of respondents	93
Average	7.258
Variant	3.568
Deviation standard	1.889
Declivity of distribution scores	-2.011
Top of distribution scores	5.054
Low score	0.000
High score	10.000
Median	8.000
Reliability coefficients	0.671
Measurements error	1.083
The average level of difficulty	0.726
The average of differences	0.549
The average differences in biseral correlation	0.802

From the results of Table.1 showed that the average item of difficulty for all items is 72 % of student. The item of difficulty in the multiple choices is medium category. From the average of differences is 54 % which is included in the category of good. The alpha value of the results of this analysis demonstrates at 0.671 that these tests have low reliability.

The results of the test analysis in terms of difficulty level, differences, and the effectiveness of the program ITEMAN 3.00 version is presented as follows:

Table 2. Reading question characteristic (*multiple choice*)

Number	Difficulty level	Category	Discriminating item	Category
1	0,914	Easy	0,651	Good
2	0,946	Easy	0,638	Good
3	0,796	Medium	0,535	Good
4	0,957	Easy	0,674	Good
5	0,914	Easy	0,590	Good
6	0,817	Medium	0,698	Good
7	0,215	Hard	0,413	Good
8	0,753	Medium	0,553	Good
9	0,505	Medium	0,443	Good
10	0,441	Medium	0,291	Fair

From the table above, results of analysis in the form of multiple choices, viewed of difficulty is: 40 % is easier, 50 % medium and 10 % is hard. The analysis showed that 90% of the reading test is accepted from the item of difficulty and discriminating power characteristics. In the difficulty level characteristics indicate that the item is still considered of measuring low-level thinking skills (Lower order thinking). While, in the differences about 10 % is sufficient. It means that the items has not been able to know the ability of students have high ability and low ability.

Table 3. The effectiveness of *Destructor* Items in Reading Test

Distraktor yang berfungsi	Questions	Number	Definition
4	6,7,9,10	4 (40%)	Good
3	3	1 (10%)	Good
2	1, 5,8	3 (30%)	Less
1	2, 4	2 (20%)	Less

From the results of Table 3 indicated that the items which has a good result is 50% and less 50 %. Therefore, 50 % of these needs to be revised to obtain a quality that can outwit the answer.

Here are the results of the manual analysis for the item in true-false as follows:

Table 4. Reading Questions Characteristics (*true false*)

No.	Difficulties	Category	Differences	Category
1	0,868	Easy	0,42	Excellent
2	0,242	Hard	0,28	Fair
3	0,670	Medium	0,33	Good
4	0,384	Easy	0,56	Excellent
5	0,285	Hard	0,36	Good
6	0,131	Hard	-0,19	Less
7	0,560	Medium	0,45	Good
8	0,428	Medium	-0,36	Less
9	0,758	Easy	0,35	Good
10	0,659	Medium	0,58	Excellent

From the results above data showed that the difficulty item of 20 % is easy, 40 % is medium and 40 % is difficult. For the index of discriminating power, 30% is good, while 50 % is medium and 20 % is less. Viewed from the item of difficulty and discriminating power, the characteristic of item were received 3, 7 and 10. The alpha value of the results of this analysis demonstrates at 0.252 that these tests have low reliability.

From the data analysis of multiple choice and true false items, revision is needed. For the reason of the results, the analysis for multiple choice tests is easy. The students in high and low ability can answer questions and the outwit item has no function up to 50%. In the true false questions, the comparison of answer key is not comparable, and leads the low quality of the item. The analysis showed 40 % of items and only three of it that good characteristic in discriminating power, so the items can showed the different of students in high and low ability.

IV. CONCLUSIONS AND SUGGESTION

A. CONCLUSIONS

1. The instrument of multiple choice has a relatively low reliability of the test is 0.671. However, the item true false showed reliability ...
2. The results showed the characteristics of multiple-choice questions for the item difficulty ranged from 0.215 up to 0.957, discriminating power ranged from 0.291 up to 0.698. 82% of multiple choice questions are accepted.
3. The result showed the characteristics of true false questions used manual Microsoft Excel, showed in discriminating power ranged from -0.19 to 0.42. In the item difficulty ranged from 0.131 to 0.868. The false true questions are 63% accepted, viewed in the item of difficulty and discriminating power.

B. SUGGESTION

Based on the analysis results of the reading test, the researcher gives the following advice:

1. Preparation of the reading test is appropriate with the standard of writing which has validation and trial so aimlessly, test qualified and eligible for use in the test.
2. Increasing the creativity of educators in arranging the instrument to measure high up to low ability

REFERENCES

- [1] American Educational Research Association, American Psychological Association, & National Council on Measurement in education (1999). *Standards of educational and Psychological testing*. Washington, DC
- [2] Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta : Mitra Cendekia
- [3] Naughton, Rose mary. 1996. *Multiple Readings in Multiple Choice Reading Test. A study of Year 11 Students' reading practice of A Multiple Choice Reading Test*, Edit Cowan University.
- [4] Reynolds, Livingston & Willson. (2008). *Measurement and Assessment in Education*. Person International edition. United States Canada
- [5] Sumadi Suryabrata. (2005). *Pengembangan alat ukur psikologis*. Yogyakarta : Andi Offset
- [6] Syaifudin Azwar. (2010). *Tes prestasi (Fungsi pengembangan pengukuran prestasi belajar)* Yogyakarta : Pustaka Belajar.

LEARNING EVALUATION MODEL DESIGN WITH MULTIPLE CHOICE TESTS FOR FIELD STUDIES EXACT SCIENCES

Nyenyep Sriwardani

Mechanical Engineering Education Department, Teacher Training and Education
Faculty, Sebelas Maret University, Surakarta, Indonesia
Jl. Ahmad Yani 200, Pabelan, Surakarta, +62 (271) 57162
daniptm@yahoo.com

Abstract - Areas of study exact sciences are subjects that have applicative characteristics are supported by the understanding and memory. In general, the evaluation is done by a subjective test, where students are given a case to be resolved or design projects that apply the formulas studied. However, with this evaluation model can not describe the state of the learner as a whole. This study will use the model to evaluate multiple-choice test on a field study to determine the exact intelligence, attitude, personality, social feeling, accuracy, and so forth. Scores processed without applying a system of penalties. Multiple choice test consisting of a question or an incomplete statement that requires an answer. To complete it must choose one of several possible answers. A selection of answers provided just the right one, while the other answer is a destructor. For students who answer inconsequential, the expression would seem when the execution of tests. This model test requires high concentration and understanding to be able answer correctly. The formulas that used for calculation must be remembered correctly and used appropriately. Therefore, it needs special techniques in practice. Design evaluation model multiple choice tests, for a field of study exact sciences, requires the preparation of a more mature both the evaluators and learners were tested. The advantages of this method are, get a test result that is objective.

Keywords: science evaluation, multiple choice test, assessment design.

I. INTRODUCTION

Education is a process of transformation of knowledge, skills and personality. In order to get the product / the maximum output, the provision of education requires careful planning in the input field, learning materials and curriculum, teacher / lecturer, teaching methods, the means (tools / media) learning, conditioning the environment (both human and non- human). To measure the success rate of the overall learning process of the evaluation program. While to assess the success of teaching a subject it is necessary to evaluate the mastery of the material. So that the results of the evaluation can demonstrate the ability learners are valid or metered correctly, it must be well planned and appropriately. The accurately assessing is one of the challenges of educational [1]. To get the exact method required an observation and research. The study can help decisionmakers, educators, researchers and educators for determining right integration technology assessment tool [1]. This research aims to design an evaluation design with multiple coice test methods in science subjects. This is a simple business to optimize the implementation of the assessment. There are a number of key drivers for conducting program evaluations Including the ability to use evaluative information [2].

A. Literature review

There are several methods of evaluation applied to a variety of different departments has proven to be beneficial to the improvement of products graduates. For example, in a study (Hoekstra, 2015) with the title of design, implementation, and evaluation of an e-portfolio approach to support faculty development in vocational education, proving that the e-portfolio potential in supporting the increase of awareness of areas of improvement, and encourage more explicit thought about planning and monitoring of professional development [3]. While the application of methods Group Education Session (GES), which is a mandatory license renewal

program for drivers aged 80 and older, in the year 1996 in Ontario. It had a positive impact on reviews their first pass attempt of the knowledge test or not [4]. The other is a research-based evaluation with the framework both cross-sectional and longitudinal de-signs [5]. This research inform the changes in the development of sports science undergraduate education in Ethiopia.

II. RESEARCH METHOD

Referring to the opinion of Clifford Woody in Silalahi, 2015 [6]. States that: research and redefining comprises defining problems, formulating hypotheses or suggested solutions; collecting, organizing and evaluating the data; conclusions making deductions and reaching either in the form of solutions towards the concerned problem or in certain generalisations for some theoretical formulation. The research method used in this research is to use predictive design fundamentally explain the phenomenon that occurred in the past and the present then analyzing situations that will be used as guidelines to create a design that will be applied in the future. In this case the value of academic achievement data analysis of the subjects of science, the mapping of the formulation of evaluation questions and analysis to perform evaluations. Then the data analysis will be used to make the evaluation design that is assumed to be successful in the future.

III. RESEARCH FINDING

The preparation of the test there are four things to be noted: the relationship between material, specific instructional objectives, learning activities and types of evaluation. Learning activities are assumed to have been carried out according to the standard and the type of evaluation is determined by the method of multiple-choice test. While the relationship between matter and aspects that are measured can be seen from the table below.

Table 1. Specifications aspects as measured on each items

Aspect Subject	Remembrance (R)	Understanding (U)	Application (A)	Total
Chapter 1 60%	4, 5, 6, 8, 9, 10, 12, 14	1, 2, 3, 7, 11, 13	15	15
Chapter 2 40%	2, 3, 4, 7, 10	1, 5, 6	8, 9	10
Total = 100%	13	9	3	25

In Table 1. It can be seen that there are two chapters of material with a theme that is still associated with the amount of about 60% to chapter 1 and 40% for chapter 2. The nature of the material in the evaluation aspect is dominated by memory, then the understanding and the final application. Each is 13 items (52%), 9 items (36%) and 3 items (12%).

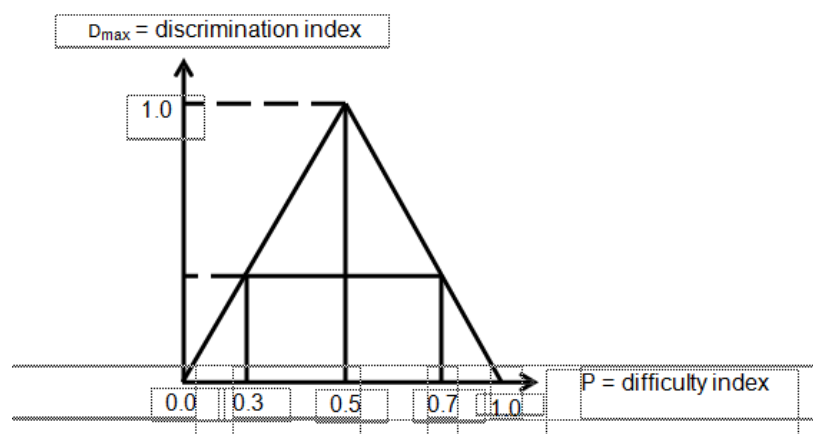


Figure 1. Index difficulty and discrimination index

Figure 1 is an illustration of the difficulty index and discrimination index. Each items will generate different charts. The problems that are too easy or too difficult to be having low distinguishing, while the matter with moderate difficulty level, will have a high distinguishing. The problem tested were taken from a standard problem with the level of difficulty is considered medium or between 0.3 up to 0.7 and with distinguishing features (discrimination index).

This study will use the models to evaluate multiple-choice test on a field study to determine the exact intelligence, attitude, personality, social feeling, accuracy, and so forth. The scores processed without applying a system of penalties. Multiple choice test consisting of a question or an incomplete statement that requires an answer. To complete it must choose one of several possible answers. A selection of answers provided just the right one, while the other answer is a destructor. This model of test requires high concentration and understanding to be able correctly answer. The formulas that used for calculation must be remembered correctly and used appropriately.

The results showed that about a successful answer is the 21-point maximum albeit with sufficient processing time span. This proves that the evaluation method is not optimal. It required an evaluation design appropriate to the character of the material science.

IV. DISCUSSIONS

Because the science subject or materials requires aspects of memory and at the same time also need to be supported by the aspect of understanding to apply, the learners will be trouble if to memorize or understand only. Trials for connecting effectiveness of idea generation methods to cognitive processes were not convincing. Engineers were asked to evaluate design ideas [7]. Discipline specific problems related to engineering processes and practices [8].

To be able to conduct an objective assessment of the capability of the application will require a special design on the conduct of the evaluation. Assessment is planned to add a pre-assessment. The expected target is first to understand the whole matter, then memorize the parts that need to be memorized, then ready to perform objective tests.

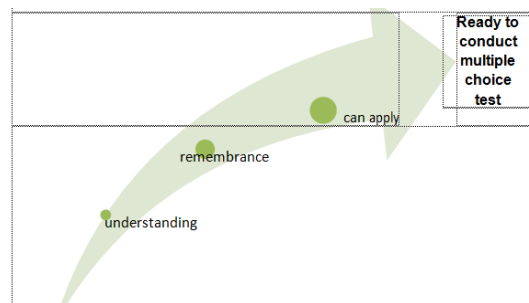


Figure 2. Pre-assessment on science subject

Figure 2. shows the priority order from smallest to largest, to the implementation of the test. So the implementation of the test is divided into 4 stages. While the evaluation value obtained from the results of the last test are multiple choice test.

- Test 1: Learners are asked to do the test questions were given the freedom to open the book and discuss (not a multiple-choice test).
- Test 2: Learners are asked to do without opening the test item reference, then given time to make a correction on the wrong parts (not a multiple-choice test).
- Test 3: Learners are asked to do without opening the application test item reference (application test).
- Tes 4: The real test (multiple-choice test).

In test 1, test 2 and test 3, the evaluation is done by a subjective test, where students are given a case to be resolved or design projects that apply the formulas studied. However, with this evaluation models can not describe the state of the learner as a whole. In its implementation to save the time, the delivery of materials to do Including at such test models one, two and three.

REFERENCES

- [1] M. Semih Summak, Mustafa Samancıoğlu, Murat Bağlibel.(2010). Technology integration and assesment in educational settings. *Procedia Social and Behavioral Sciences*, vol. 2, pp. 1725–1729.
- [2] Michelle Ardisson, Dnp. Rn. Acnp-Bc, Benjamin Smallheer, Phd. Rn. Acnp-Bc, Ccrn, Ginny Moore, Dnp. Rn. Aprn-Bc, And Tom Christenbery, Phd. Rn. Cne. *Meta-Evaluation: Experiences In An Accelerated Graduate Nurse Education Program*. Science Direct.

-
- [3] Annemarieke Hoekstra, Jocelyn R. Crocker. (2015). Design, implementation, and evaluation of an ePortfolio approach to support faculty development in vocational education. *Studies in Educational Evaluation*, vol. 46, pp. 61–73.
 - [4] Ward Vanlaar, Marisela Mainegra Hing, Robyn Robertson, Dan Mayhew, David Carr. (2016). An evaluation of Ontario's Group Education Session (GES) for license renewal of seniors aged 80 and above. *Journal of Safety Research*, vol. 56, pp. 33-39.
 - [5] TeferaTadesse, SisayMengistu, YehwualashetGorfu. (2016). Practice papers: Using research-basedevaluationtoinformchangesinthe development of undergraduate sports science education in Ethiopia. *Journal of Hospitality, Leisure, Sport & Tourism Education*, vol. 18, pp. 42-50.
 - [6] Ulber Silalahi. (2015). *Metode Penelitian Sosial Kuantitatif*. ISBN 978-602-7948-78-5. Bandung: PT Refika Aditama.
 - [7] Stéphanie Fischera, David Ogetb, Denis CavalluccibaThéodile. (2016). The evaluation of creativity from the perspective of subjectmatter and training in higher education: Issues, constraintsand limitations, vol. 19, pp. 123-135.
 - [8] Jurgis Kazimieras Staniškis, Egl é Katiliūt é. (2016). Complex evaluation of sustainability in engineering education: case & analysis. *Journal of Cleaner Production*, vol. 120, pp. 13-20.

BHAGAVAD GITA VIDEO FOR HINDUISM EDUCATION LAMPUNG

Nyoman Siti, I Komang Arteyasa, Ni Made Indrayani
Mahasiswa Pasca Sarjana, Universitas Lampung
IT Department, Sekolah Tinggi Agama Hindu Lampung
Finance Department, Sekolah Tinggi Agama Hindu Lampung
nyomansiti123@gmail.com

Abstract - This research is a developmental research. This research aims to determine the effect of video media Bhagavad Gita to the learning outcomes of students and the effectiveness of the media for independent study. This research method using a hypothetical model library method is done in three stages: (1) a preliminary study, (2) the development of the prototype, (3) output the results in the form of a hypothetical final. The results of this study are: first, the use of video media in the learning process improve their learning outcomes. This shows that the video Bhagavad Gita can enhance children's learning interpretation. second, students' learning behavior after using the medium of video tutorials are (1) Students are very interested in using the medium of video tutorials. This can be seen in all of the students who look to focus watched a video tutorial on their individual computers. (2) The students seemed excited and (3) Students are seen actively want to quickly witnessed and experienced the learning event using video tutorial. (4) Students are also seen to be proactive to participate practice step by step tutorial to get the video. (5) All students have no difficulty in running a video tutorial. (6) Students are visible passion for performance capabilities appropriate practice as exemplified in the video tutorial. (7) Students do not seem desperate at the time of practice, they failed or did not correspond to what is exemplified in the video tutorial. They continued to try to repeat. Repeating watching video tutorials and also repeat performance. The students look independence. (8) Students are savvy enough to use this learning media.

Keywords: *Learning Outcomes, Video Bhagavad Gita*

I. INTRODUCTION

Education is a conscious and deliberate effort to create an atmosphere of learning and the learning process so that learners are actively developing the potential for him to have the spiritual power of religion, self-control, personality, intelligence, character, and skills needed him, the community, the nation and the State (Depdiknas: 2003, Soekidjo Notoatmodjo: 2003). Improving the quality of education is directed to the development of science, technology, arts oriented to the needs of learners.

Learning is a process of change in a personal behavior caused by experience and training (Hamalik: 2009, Sugandi: 2000, Slameto: 2003). Learning is a system that aims to help the students learning process, contains a series of events designed, arranged in such a way to influence and support the students learning process which is internal in an effort to teach students (Uno: 2009, Gagne and Briggs: 1979). Required media in learning to facilitate learners.

Instructional media are materials, tools, or techniques used in learning activities with the intention that the process of communication interactions education between teachers and students may take appropriate effective and efficient in stimulating learning (Latuhu: 1988, AECT: 1977, Rival: 1978). The use of media as a tool not only able to attract students to the learning process, but also helps increase understanding of learners and facilitate educators convey science (Darvina: 2003). Based on these definitions, instructional media have a great advantage in facilitating students to learn the subject matter. Learning media can be classified into several types. Based on sound elements, namely: complex media (film, television, video / VCD) and simple media (slides, audio, transparency, text). Also the media are also classified by

its scope, ie media mass (coverage of comprehensive and simultaneous / radio, television), media group (the coverage area of the room / cassette audio, video, OHP, slide, etc.) and media individually (for individuals / book text, phone, CAI). With so many media are then expected to facilitate the delivery of information in the learning process (Schramm: 1985 Heinrich, Molenda, Russell: 1996).

Media video is a series of electronic images accompanied by audio sound elements, it also has a graphic element that is poured through a video tape (Ronald Anderson: 1994). Very helpful instructional media facilitate learners in learning and make the learning process more interactive, effective, efficient and attractive (Andi Kris: 2011, Nadya Putri: 2012). Then the teacher or instructor advised to use video as an alternative to learning, in order to get better learning outcomes for more attractive (Setiawan Hendarto: 2012, Siti Zubaidah, Mujiyono Wiryotinoyo, Sudaryono: 2014).

Bhagawadgita (Sanskrit: भगवद्गीता; Bhagavad-Gita) is a famous part of the Mahabharata, in the form of a dialogue that is poured in the form of a poem. In this dialogue, Krishna, the Supreme Personality of Godhead is the main speaker outlining the teachings of Vedanta philosophy, while Arjuna, a direct disciple Sri Krishna who became listeners. Literally, the meaning of Bhagavad-gita is "Songs of Sri Bhagavan (Bhaga = greatness perfect, van = has, Bhagavan = Who has the power of perfect, good looks perfect, unlimited wealth, fame eternal, infinite power, intelligence infinite, and a perfect detachment, owned simultaneously).

This verse is an interpolation or inserts to the "Dragon". This scene took place at the start side, or the war in Kurukshetra. At that time Arjuna standing in the middle of the battlefield of Kurukshetra between the Kauravas and the Pandavas. Arjuna worried and hesitant because that will be fought war were relatives, friends and teachers. Then Arjuna granted true knowledge about the secret of life (spiritual) is Bhagawadgita by Krishna Arjuna accepted as a coachman at the time (Wikipedia: 2015, Pudja: 2005)

Based on these exposures indicates that the current learning must use a variety of media such as video lessons, but on the field there are teachers who do not use it, especially teachers of Hinduism. Thus, the need for the development of the management model of learning resources in the learning Hinduism

II. RESEARCH METHODOLOGY

This paper analyzed using library method hypothetical model that research is done in three stages: (1) a preliminary study, the assessment of the condition management of learning resources factual / existing and literature studies, (2) development of a prototype hypothetical model management based learning resources character, and (3) outcomes of the research, which produces a final hypothetical model. Here is a picture frame of the study.

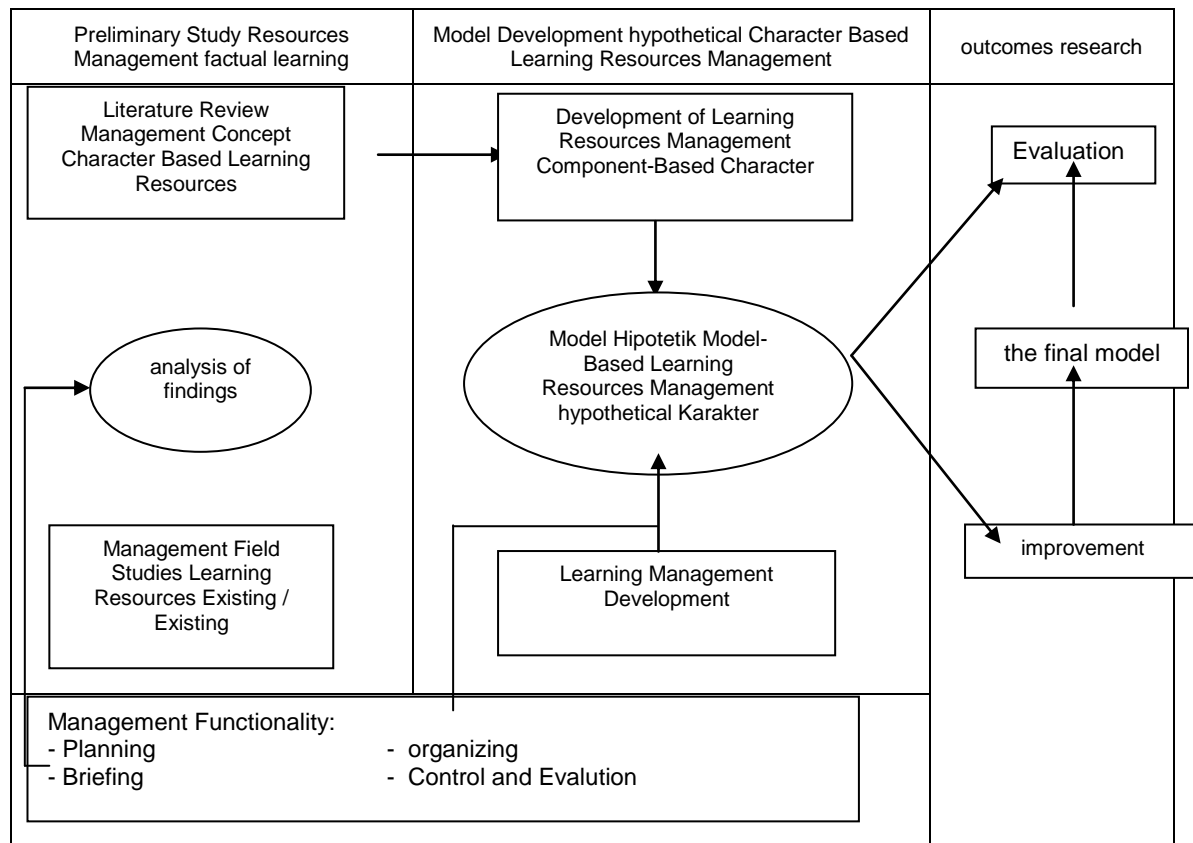


Figure 1. Frame of the research

III. RESULT AND DISCUSSION

Development of instructional video media is necessary, given the instructional video media in fact be petrified teaching and learning process in the delivery of the subject matter so that the material can be conveyed more concrete with a high degree of realism. Media video that can be heard or seen to make it easier to reconstruct the student learning experience. In addition, the development of video media also solves the problem of learning is hampered because of time, cost and space. Tenny Widya Kristiana (2012)

The role of media in teaching and learning by Gerlac and Ely (1971: 285) stated that there are three privileges owned media, namely: (1) Media has the ability to capture, store and display the back of an object or event, (2) Media has the ability to redisplay the objects or events in various ways tailored to the needs, and (3) Media has the ability to display an object or event that implies.

Student achievement has increased significantly with an animated video based on a comparison of the experimental group to the control group. Learning media by using animated video shown to contribute to the improvement of the quality of student learning, but the inside is still less than perfect. The use of instructional media can enhance students' understanding, it should be for the lessons that are applicable in the use of instructional media to help students understand the material given by the teacher. (Setiawan Hendarto, Sunyoto & Widya Aryadi: 2012)

Observation of student learning behavior after using the medium of video tutorials are (1) Students are very interested in using the medium of video tutorials. This can be seen in all of the students who focused watched a video tutorial on their individual computers. (2) The students seemed excited and (3) Students are seen actively want to quickly witnessed and experienced the learning event using vdeo tutorial. (4) Students are also seen to be proactive to participate practice step by step tutorial to get the video. (5) All students have no difficulty in running a video tutorial. (6) Students are visible passion for performance capabilities appropriate practice as exemplified in the video tutorial. (7) Students do not seem desperate at the time of practice, they failed or did not correspond to what is exemplified in the video tutorial. They continued to try to

repeat. Repeating watching video tutorials and also repeat performance. The students look independence. (8) Students are savvy enough to use this learning media. They were able to learn independently. Teachers simply provide motivation and assistance, if any students would like to ask. The role of teachers in teaching replaced by media. Revi, Syahvani, Dede (2015)

Concluding the interview by Revi, Syahvani, Dede (2015) to 22 students after using the video tutorial is (1). Students argue, learning by using video media tutorial very well be applied to the material administer the server in the network. (2). Students in the class say no difficulty in the operation of media video tutorials (3). Media video tutorials greatly assist students in understanding the material. In addition to practice, there is also the theory of a preliminary order by the students as well, mature concept. (4). Students can understand the material presented through computer network. The material is presented in a coherent and focused. (5). Most of the students stated that by using the tutorial video media, students are able to study independently, not dependent on the presence of a teacher.

IV. CONCLUSION

Based on the description above, it can be deduced which is as follows: first, student achievement has increased significantly with an animated video based on a comparison of the experimental group to the control group. This shows that the video Bhagavad Gita can enhance children's learning interpretation. Second, students' learning behavior after using the medium of video tutorials are (1) Students are very interested in using the medium of video tutorials. This can be seen in all of the students who look to focus watched a video tutorial on their individual computers. (2) The students seemed excited and (3) Students are seen actively want to quickly witnessed and experienced the learning event using video tutorial. (4) Students are also seen to be proactive to participate practice step by step tutorial to get the video. (5) All students have no difficulty in running a video tutorial. (6) Students are visible passion for performance capabilities appropriate practice as exemplified in the video tutorial. (7) Students do not seem desperate at the time of practice, they failed or did not correspond to what is exemplified in the video tutorial. They continued to try to repeat. Repeating watching video tutorials and also repeat performance. The students look independence. (8) Students are savvy enough to use this learning media.

V. SUGGESTION

Based on these conclusions, the researchers, who will conduct similar studies recommended in order to make video Bhagavad Gita to enhance and simplify the learning more effective and easier for students to learn the any scriptures.

REFERENCES

- [1] Anderson, H. Ronald, *Pemilihan dan pengembangan media Video pembelajaran*, Grafindo Pers, Jakarta, 1994.
- [2] Darvina Omar, *Kajian pembangunan dan penilaian bahan pembelajaran berbantuan komputer berasaskan cd interaktif bagi mata pelajaran pemasaran 1 di Politeknik*, Tesis Sarjana, Kolej Universiti Teknologi Tun Hussein Onn, 2003.
- [3] Depdiknas, *Undang-undang Republik Indonesia No. 20 tahun 2003 tentang sistem pendidikan nasional*, Sekretaris Negara RI, Jakarta, 2003.
- [4] Ely, Gerlac, *Teaching and media systematic approach*. New Jersey Prentice Hall, Inc, 1971.
- [5] Gagne, M. Robert dan J. Briggs Leslie J, *Principles of instructional design*, New York: Rinehart and Winston, 1979.
- [6] Hamalik, Oemar, *Media pendidikan*, Bandung : Penerbit Alumni, 1980.
- [7] Heinich, Robert, Michael Molenda, dan D. James Russell, *Instructional media and the new technologies of instruction*, United State of Amerika: SAGE Publications, Inc, 1996.
- [8] A. Kristanto, *Pengembangan model media video pembelajaran mata kuliah pengembangan media video/tv Program Studi Teknologi Pendidikan Fakultas Ilmu Pendidikan Universitas Negeri Surabaya*. Jurnal teknologi pendidikan, 2011, 11(1), 12-22, diakses tanggal 21 April 2014
- [9] T. W. KRISTIANA, *Pengembangan media video pembelajaran pertempuran di surabaya siswa kelas ix SMP Negeri 1 Kalitidu-Bojonegoro*, SKRIPSI Jurusan Sejarah-Fakultas Ilmu Sosial UM, 2012.
- [10] Lathuru, D. Jhon, *Media pembelajaran dalam proses belajar mengajar masa kini*, Jakarta: P2LPTK, 1988.
- [11] Pudja, Gede, *Bhagawad Gita (Pancama Weda)*. Surabaya: Paramita, 2005.
- [12] N. Putri, *Efektifitas penggunaan media video untuk meningkatkan pengenalan alat musik daerah pada pembelajaran ips bagi anak tunagrahita ringan di SDLB 20 Kota Solok*, Jurnal ilmiah pendidikan khusus, 2012, 1(2), 318-328.
- [13] A. Rivai, *Apa dan mengapa media pendidikan*, Bandung: LPP BPP IKIP Bandung, 1978.
- [14] R. R. Saleh, D. Ramdani, *Pengembangan media video tutorial dalam pembelajaran komputer untuk keterampilan membuat server di smk*, Jurnal pendidikan dan pembelajaran, 2015, 4(10).
- [15] Schramm, Wilbur, *The process and effect of mass communication*, New York: University Of Illinois Press, 1985.

- [16] D. Setyawan, Pengembangan media audio video untuk meningkatkan hasil belajar pada siswa SMK PN2 Purworejo, AUTOTECH-pendidikan teknik otomotif, 2015, 5(1).
- [17] Slameto, Belajar dan faktor-faktor yang mempengaruhinya, Jakarta: Rineka Cipta, 2003.
- [18] Soekidjo Notoatmodjo, Pendidikan dan perilaku kesehatan. Jakarta : PT Rineka Cipta, 2003.
- [19] Sugandi, Achmad, dkk, Teori pembelajaran, Semarang: UPT MKK UNNES, 2004.
- [20] S. Zubaidah, M. Wiryotinoyo & S. Sudaryono, Pengembangan video pembelajaran menulis puisi untuk siswa kelas v sekolah dasar, Jurnal Tekno-pedagogi, 2014, 4(1).
- [21] <https://id.wikipedia.org/wiki/Bhagawadgita> diakses tanggal 13 September 2015

DEVELOPMENT OF AUTHENTIC ASSESSMENT INSTRUMENT AT GRADE FOUR ELEMENTARY SCHOOL IN MALANG

Puri Selfi Cholifah¹⁾, Muhardjito²⁾, Eddy Sutadji³⁾

¹ Elementary Education, State University of Malang

² Physics, State University of Malang

³ Mechanical Engineering, State University of Malang

¹ selfi.cholifah@gmail.com

Abstract - This research is aimed at producing a product in the form of an authentic assessment at grade four in Malang. The model of the research and development in this study refers to the ADDIE development model and assessment development model O'Malley and Pierce. The subjects of tryout in this study include assessment experts, linguists, and that teachers and field practitioners of fourth grade students of SDN Bunulrejo 2 Malang. The results showed that the level of validation are at a very valid level with a percentage of 96.67%. Based on the test results, the statistical computation used was Spearman correlation coefficient, the data was obtained from the inter-rater reliability (IRR) results of correlation test using Spearman that was stated as 56 descriptors declared reliable and 2 descriptors obtained as unreliable. Based on the inter-rater agreement (IRA) of the Kappa statistical test that is interpreted of Landis and Koch, there were two categories rubric that were almost perfect (> 0.8), eight rubrics were substantial (0.61 to 0.8) and four sections that were still in moderate (0.41 to 0.6). In general, the practical score was equal to 94% with predicate very practical and also qualitative data in the form of the interview so that it can be said as good. This statement is in line with the results of the questionnaire of the teachers that reached 93.3%. Based on the results, it was obtained some revisions to enhance both products including the determination of essential indicators rubric and variant tasks.

Keywords: *instrument, assessment, authentic*

I. INTRODUCTION

The change of curriculum in Indonesia is closely related to the quality enhancement of education. One critical point in implementing 2013 curriculum is that the authentic assessment as an integral part of thematic learning. In an integrative thematic learning, the evaluation made is based on authentic assessment [1]. This causes a shifting paradigm in assessment that is change or reformation in the field of curriculum, teaching, and assessment to prepare students for the life and working context of 21st century [2].

The form of authentic assessment might be various, however the determination of whether it is authentic or not is based on the provision of tasks which are relevant to the real condition experienced by students. The general forms of authentic assessment can be conducted through project assessment, portfolio assessment, performance assessment, self- assessment, and peer-assessment.

Performance assessment can be defined as one way or method to evaluate knowledge, concept, and skill by obligating students to perform their task and simulate the real condition in applying certain knowledge, concept or skill [3,4]. Portfolio is one clustered sample of task with a collection process in a certain range of time [5,6,7,8]. Portfolio assessment is one part of authentic assessment with a data collection stage through collecting development process or students' product in a certain time. Project assessment deals with assessment conducted to finish a task in a certain time. The task can be in the form of investigation from the planning, data collection, as well as data processing and presenting [9].

The aforementioned assessments are included in authentic assessment in case the implementation is directed into students' real life. Thus, appropriate mastery of assessment and assignment as well as assessment development containing criteria of valid, reliable, and practical allow teachers to minimize subjectivity and result data describing the real condition of students aimed in line with the objective of assessment conducted.

Validity is related to a tool for measuring a particular object. Concept of validity is about the precision of a tool to measure an object that must be measured [10,11,12]. In the development of authentic assessment, the validity included is content validity. Content validity is related to how far an assessment represent a domain that will be measured [6,13].

The next concept in authentic assessment instrument is that reliability. Reliability deals with consistency, stability, and dependability of an instrument [6,11,14]. The use of inter-rater reliability test is required in case one thing that will be found out is the similarity index between two or more correctors [13].

A quality assessment instrument must be practical and efficient. Some components influencing the practicality of an assessment instrument are: (a) teacher must be accustomed to any method used, (b) time required, (c) administrative complexity, (d) ease of scoring, and (e) ease of data interpretation [6].

This study is aimed at resulting a product in the form of instrument of authentic assessment for IV graders of Elementary School in Malang. Product of authentic assessment developed in this study has some aspects: valid, reliable, practical, and effective. Instrument specification of authentic assessment developed consists of book of rubric instrument assessment and guidance. The rubric of authentic assessment contains an analytical rubric with scale of 4 intentionally presented to teacher in conducting authentic assessment. Rubric of authentic assessment instrument developed includes instrument of performance assessment, portfolio assessment, and project assessment. The second product is the guidance book of authentic assessment conduct becoming a guidance for teacher in the application of instrument being developed.

II. METHOD

The model of research and development in this present study refers to the procedural model by following the R&D ADDIE and assessment development model of O'Malley & Pierce since both complement each other in the required process. The stage includes analyzing, designing, development, implementation, and evaluation. While the stage of instrument development of assessment which has been modified has been set including determining the purpose, specifying the objective, conducting staff development, collecting sample assessment, adapting existing assessment, trying out the assessment and reviewing the assessment.

The participant involved in this research and development consisted of assessment expert, linguist, and field practitioner (teacher) as well as IV graders of SDN Bunulrejo 2 Malang in many of 28 students. Assessment expert and linguist dealt with aspect of validity criteria of the developed product. Teacher of IV grader subject presented data on reliability, practicality, and effectiveness of development product being conducted. The corrector involved consisted of two people, one of them is a teacher of SDN Bunulrejo 2 Malang and the other one is a teacher from another school.

III. FINDING AND DISCUSSION

After following some stages as determined in research and development steps, then a new product in the form of authentic assessment instrument for IV graders of Elementary School in Malang was resulted. The final product resulted was in the form of performance assessment, project assessment, and portfolio assessment in the form of rubric with 4 criteria: validity, reliability, practicality, and effectiveness.

A. Validity

Validity test conducted is validity test of assessment expert and that of linguist. Validity which is studied is content validity in conformity with curriculum as the purpose of assessment having direct relation to the learning objective as the deviation of standard content (SI) and basic competence (KD). In general, those viewed from authentic assessment instrument is that in the form of content validity or curriculum validity [15]. Through this validity, an optimism is gained to conduct assessment on thinking skill which is in line with the curriculum being implemented. The

significant aspect of authentic assessment at least must include: problem solving skill and high order thinking skill [16].

Data from validation result shows that rate of content validation or the conformity of instrument with assessment expert was considered very valid with percentage of 96.67%. In terms of language aspect, Validator also gave a high percentage that as 83.29% which was considered to be very valid. The two data were considered very valid with validity value >75% if it is consulted.

The rate of content validity in an assessment instrument depends on analysis of latticework that have been made. Some steps that can be implemented to design a test with high content validity are: (1) identifying competence in detail to get the indicator detail in accordance with the competence, (2) setting latticework completely and in detail by listing tasks, (3) elaborating and writing the tasks in accordance with the latticework [12].

B. Reliability

The data resulted from the data collection are those related to instrument reliability which was obtained after recapitulating correction result made by corrector 1 and corrector 2. Statistical test used Spearman index to obtain coefficient of inter-rater reliability (IRR). Whereas, in terms of inter-rater agreement (IRA), Cohen Kappa value were used.

There were 2 items considered had not yet met the standard of the determined reliability, they are N1d in descriptor 3 that was in use of media which has coefficient of inter-rater reliability in amount of 0.556 and rubric N3c descriptor 4 on making presentation which only reached coefficient reliability in amount of 0.578. The two items were considered to be improved in order to can be used as a more reliable assessment instrument.

The test of inter-rater agreement (IRA) was conducted by Kappa test. The result then was presented in an interpretation table of Landis and Koch. The recapitulation of data for each rubric was presented in the following table. The data resulted from the data collection are those related to instrument reliability which was obtained after recapitulating correction result made by corrector 1 and corrector 2. Statistical test used Spearman index to obtain coefficient of inter-rater reliability (IRR). Whereas, in terms of inter-rater agreement (IRA), Cohen Kappa value were used.

There were 2 items considered had not yet met the standard of the determined reliability, they are N1d in descriptor 3 that was in use of media which has coefficient of inter-rater reliability in amount of 0.556 and rubric N3c descriptor 4 on making presentation which only reached coefficient reliability in amount of 0.578. The two items were considered to be improved in order to can be used as a more reliable assessment instrument.

The test of inter-rater agreement (IRA) was conducted by Kappa test. The result then was presented in an interpretation table of Landis and Koch. The recapitulation of data for each rubric was presented in the following table:4

Table 1. Recapitulation of Inter-Rater Reliability Average (Irr) and Inter-Rater Agreement (IRA)

Rubric Code	Spearman	Kappa
N1a	0,831	0,689
N1b	0,799	0,708
N1c	0,780	0,489
N1d	0,737	0,643
N2a	0,807	0,742
N2b	0,813	0,618
N2c	0,749	0,583
N2d	0,973	0,930
N3a	0,816	0,678
N3b	0,952	0,879
N3c	0,772	0,529
N3d	0,666	0,534
F	0,841	0,790
P	0,793	0,619

Based on table 1, it was found out that inter-rater reliability of Spearman correlation rubric instrument of performance assessment, portfolio assessment, and project assessment were all considered reliable. While the agreement among raters was categorized in three parts. Two rubrics were considered has value category agreement which was almost perfect (>0.8) they were N2d and N3b, eight rubrics were considered in conformity with the category (0.61-0.8) they were N1a, N1b, N1d, N2a, N2b, N3a, F, and P as well as four rubrics were considered in medium category (0.41-0.6) they were N1c, N2c, N3c, dan N3d.

Reliability concept in authentic assessment instrument being developed consider two things: stability and consistency. Inter-rater agreement and inter-rater reliability are two separated indexes yet both of them are worth to consider [17]. Thus, inter-rater agreement is in line with stability of rater which are various for each student, while inter-rater reliability is consistency of value from corrector which is different for all student [18].

The calculation of rate of inter-rater agreement was by using formula of Copen Kappa to find out the rate of agreement of rubric used [19,20]. Elaboration of agreement rate in Kappa degree was divided into 6 categories based on benchmark Landis and Koch, they are very low (<0.0), low (0.0-0.2), medium (0.41-0.60), (0.21-0.4), appropriate 90.61-0.80) and almost perfect (>8) [21]. Whereas the correlational test used between two value of two raters (inter-rater reliability) might follow coefficient of Pearson or Spearman rho [11,18].

Data gained in this development were in rubric N1c on report presentation obtaining inter-rater reliability with coefficient of 0.818 yet it only had rate of agreement which is considered medium, 0.464. From the data, it was found that corrector gave many different score for each student in the descriptor. However, if it were reviewed in terms of its inter-rater reliability then it was found that that students that should get high score are scored high by another corrector. This showed that the rate of consistency among raters was deemed to be high even though the rate of inter-rater agreement was low.

C. Practicality

Data of practicality was obtained by reviewing the ease of using, ease of value conclusion, clarity of task description, clarity of product presentation, completeness of scoring presentation, clarity of scoring guide, clarity of assessment instrument mapping guide, clarity of assessment rubric, clarity of scoring guide, and clarity of scoring interpretation with maximum score of 4 or 100%. The next descriptor is clear and comprehensible activity guide and assessment target of guidance book obtained 3.67 or 92%. While the descriptor of ease of result interpretation, completeness of scoring procedure in guidance book, and clarity of implementation time in assignment, it was obtained score of 3.33% or 83%. In general, the supporting practicality data that is an average score, 94% is obtained with very practical predicate and also qualitative data in the form of supporting interview so it was considered good.

Administrative complexity, ease of scoring, and ease of data interpretation presented in the instrument was clearly presented so the table of consultation gave space for teacher to ease scoring without processing in a complicated calculation. In the development of authentic assessment instrument for IV graders of Elementary School in Malang. Rubric represented performance assignment performed by students. The use of analytical rubric is more appropriate to make a scoring result based on learning objective (objective-by-objective [22].

D. Effectiveness

Effectiveness data was obtained through thematic analysis by using qualitative approach. The performance result obtained in this study involves: time efficiency, in a row order scoring stage and response obtained after the conduct of authentic assessment. In data collection stage in general, teacher used to communicate the guidance, scoring and stage of conduct in accordance with assessment book. The guidance and scoring information used to be delivered at the beginning of activity. It was noted in a note field in the form of observation result in Mei 8, 9, 12, and 13 2015. This indicated that rate of effectiveness of teacher performance was considered effective. The statement was in line with the result of teacher's questionnaire reaching 93.3% obtaining very high rate of effectiveness.

An effective assessment means making assessment in more fair and efficient way as well as having contribution to learning. Effectiveness is defined as "how well something works or how intense something is in accordance with the objective being set [23]." Effectiveness data of this study were investigated in terms of performance, meeting purpose, goal achievement.

Class teacher is a user as well as implementer since assessment is one integral part that should be owned by a teacher in other words, teacher should be skillful in administrating, scoring, and interpreting data of assessment result [24]. Besides, in terms of reaching effectiveness of instrument assessment, students are required to know some aspects that will be evaluated during assessment process. This also must be conducted by teacher in order to give information of aspects that will be evaluated

E. Product Revision

Suggestion, critic, and comment from teacher then were implemented in a constructive improvement in revision stage. The revision dealt with the improvement of development product being conducted. Based on the analysis result of data and the constraint as well as from validator, teacher and investigation result of quantitative and qualitative data then a revision stage was required. The improvement can be made by changing final product of evaluation stage as well as review of the assessment result. The revised component is instrument of performance and project including rubric indicator (N1D, N2C, N3C, N2A, P2), assignment diversification, product presentation, and language aspect. Based on data of the result of data collection revision, it was obtained some revisions to improve product whether in the form of rubric, assignment variant, and scoring guidance. This was aimed at increasing the reliability coefficient and inter-rater agreement, the rate of practicality and effectiveness of authentic assessment instrument for IV grade of Elementary School in Malang.

IV. CONCLUSION

The result of this present research and development of authentic assessment instrument is in the form of valid, reliable, practical, and effective book of assessment instrument rubric. Data of validation result showed the rate of content validity or the conformity with the instrument of assessment expert in a very valid rate with percentage of 96.67%. the same thing was also given by validator in terms of language aspect with a validity percentage of 83.39% which is considered very valid. The two data are considered as very valid criteria that can be consulted with validity value of >75%. Reliability data showed that 56 descriptors have considered it as reliable (>0.6) based on the review of inter-rater agreement from statistical test of Kappa, there were two categories of rubric that were almost perfect (>0.8), eight rubrics of category were appropriate (0.61-0.8) and four rubrics were still in medium category (0.41-0.6). Practicality data obtained from interview, observation and documentation showed that the rate of effectiveness of teacher performance quality was considered effective. The result was also supported by questionnaire of effectiveness which was in percentage of 93.3%. The revision stage was conducted by perfecting the products including rubric, assignment variant, and scoring guide. This was conducted to increase validity, reliability coefficient, and inter-rater agreement, rate of practicality as well as effectiveness of the authentic assessment instrument for IV grade of Elementary School in Malang.

ACKNOWLEDGMENT

Puri Selfi Cholifah would like to say million thanks to the co-author, Eddy Sutadji and Mahardjito who always give support and improvement in the present study in terms of providing valuable feedback. Praise also be gratitude to Allah the Almighty for the successful conduct of this study as well as the stage of this publication, and the researcher would like to say thanks to any party directly or indirectly involved in this study. We will not be able to give a beneficial knowledge contribution from research result publication without the support of all parties that cannot be mentioned one by one of you all here.

REFERENCES

- [1] S. Kovalik, *Integrated Thematic Instructions: The Model*. Kent: Suvan Kovalik & Associates, 1994.
- [2] L. Darling-Hammond, "Policy frameworks for new assessment," in *Assessment and Teaching in 21st Century Skills*, P. Griffin, B. McGaw., & E. Care (Eds.), New York: Springer, 2012, pp. 301-339.
- [3] D. G. Wren, *Performance Assessment: A Key Component of Balanced Assessment System*. Virginia Beach: Department of Research, Evaluation, and Assessment, 2009.
- [4] C. Oberg, "Guiding Classroom Instruction Through Performance Assessment." *Journal of Case Studies in Accreditation and Assessment*, vol 1, September 2010, accessed August, 28 2014, retrieved from <http://www.aabri.com/manuscripts/09257.pdf>.
- [5] M. Çakan, G. Mıhladıız, & B. Göçmen-Taşkın, "How portfolio use affects students' learning and their attitudes toward 6th Grade Science Lesson". *International Online Journal of Educational Science*, vol 2 (2), pp. 362-377, 2010.

- [6] J.H. McMillan, *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*. Boston: Pearson Education, 2007
- [7] R. I. G. Lucas, "A study on portfolio assessment as an effective student self-evaluation scheme," in *The Asia Pacific-Education Researcher*, vol 16 (1) pp. 23-32, 2007.
- [8] S. Dikli, "Assessment at a distance: traditional vs. alternative assessments".*The Turkish Online Journal of Educational Technology*, vol 2 (3), pp. 13-19, July 2003, accessed August, 28 2014, retrieved from <http://www.tojet.net/articles/v2i3/232.pdf>.
- [9] H. A. Asep, *Evaluasi pembelajaran*. Yogyakarta: Multi Pressindo, 2013.
- [10] K. D. Moore, *Effective Instructional Strategies: from Theories to Practice (Fourth Edition)*. London: SAGE Publication, 2014.
- [11] R. J. Cohen & M. E. Swerdlik, *Psychological Testing and Assessment: An Introduction to Test and Measurement (Seventh Edition)*. New York: McGraw Hill, 2010.
- [12] T. Harsiati, *Penilaian dalam Pembelajaran (Aplikasi pada Pembelajaran Membaca dan Membaca)*. Malang: UM Press, 2011.
- [13] N. J. Salkind, *Test & Measurement for People Who (Think They) Hate Test & Measurement*. Thousand Oaks: SAGE Publications, 2013.
- [14] R. E. Traub, *Reliability for the Social Science: Theory and Applications*. Thousand Oaks: SAGE Publication, 1994.
- [15] J. M. O'Malley, & L. V. Pierce, *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. Addison-Wesley Company, Inc., 1996.
- [16] Ronis, D. *Asesmen Sesuai Cara Kerja Otak*. Jakarta: PT Indeks, 2011.
- [17] J. W. Fleenor, J. B. Fleenor, & W. F. Grossnickle, "Interrater reliability and agreement of performance ratings: A methodological comparison." *Journal of Business and Psychology*, vol 10 (3), pp. 367-380, 1996.
- [18] S.C. Liao, E. A. Hunt, & W. Chen, "Comparison between Inter-rater Reliability and Inter-rater Agreement in Performance Assessment." *Annals Academy of Medicine*, vol 39 (8) pp. 613-618, 2010.
- [19] M. Cunningham, "More than just the kappa coefficient: a program to fully characterize inter-rater reliability between two raters". Paper presented at SAS Global Forum Statistic Data Analysis. Paper 242, 2009.
- [20] K. A. Hallgren, "Computing inter-rater reliability for observational data: an overview and tutorial." *Tutorials in Quantitative Methods for Psychology* vol 8(1), pp. 23-34, 2012.
- [21] K. L. Gwet, *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring The Extent of Agreement Among Raters*. Gaithersburg, MD: Advanced Analytics, LLC, 2012.
- [22] R. J. Marzano, D. Pickering, & J. McTighe, *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Alexandria: Association for Supervision and Curriculum Development, 1993.
- [23] C. Chanchitpricha, "SEA effectiveness criteria for measuring impact assessment tools implementation and practice". Paper presented at Making an Impact? IAIA Proceedings, di Prague, Republik Czech, 21-23 September 2011.
- [24] G. K. Cunningham, *Assessment In The Classroom: Constructing and Interpreting Tests*. London: The Falmer Press, 1998.

MODEL EMPLOYEE PERFORMANCE EVALUATION OF ECONOMICS GRADUATE DEGREE IN BALI

Putri Anggreni

Universitas Mahendradatta, HEPI Pusat

e-mail: gekcay@gmail.com

Abstract - This study aims to: (1) develop standards of employee performance graduate degree in economics; (2) develop instruments to measure employee performance; (3) describe the profile of employee performance; (4) evaluating the performance of employees graduate degree in economics; (5) get an effective model to employee performance graduate degree in economics. This study consisted of two phases: (1) the development stage of employee performance standard economics graduate and development of instruments for measuring employee performance; (2) research evaluation. The sampling technique used is proportionate stratified random sampling. The subjects were 1,000 employees economics graduates in Bali. The construct validity of the grains questionnaire determined by factor analysis and reliability determined by the reliability index based on Cronbach alpha formula. The reliability index for instrument items is also measured by means of inter-rater. The results showed that: (1) the indicators used to measure the performance of employees workplace, job skills, motivation and personality. Based on hthe analysis of grains instrument can be concluded that by using measures of employee performance measurement instrument can develop economics graduates expressed in this research can yield precise measurement instrument, which has proven the validity, reliability, and accuracy. The results of the evaluation includes four levels namely: (1) the level of learning; (2) the level of behavior; (3) the level of results; (4) the level of reaction.

Keywords: *economics graduate, evaluation model, employee performance.*

I. INTRODUCTION

The opening of the vast employment opportunities in the economic field, does not mean it can automatically absorb the labor force because there is a gap (mismatch) between the qualifications required by the employment opportunities available competence. Entering the MEA, will flood also labor from ASEAN countries so that there will be a very tight competition. The main key to win this competition by improving the quality of human resources. Indonesia, as predicted by the World Bank, will be one of the top five economies of the XXI century. This potential can only be realized if the human resources prepared to face the competition in the industry work.

One way in which the industry is now working to obtain quality human resources and has a high work professionalism is through a system of recruitment is done.

In the S1 level, recruitment of graduates are the most common activities performed by the existing work industry. The main determinants that influence the selection of employment to graduates of S1 Faculty of Economics are: (1) The terms of office requested; (2) Work experience; (3) The organization's budget constraint; (4) Fees for talented people who will be recruited (salary); (5) Market competition; and (6) Reputation almamater.

As a small island, Bali does have limited carrying capacity, both physically and carrying capacity of the environment as a whole. Therefore, the development of Bali to be carefully considered so that sustainability can be implemented, particularly in the provision of quality human resources and professionals in the field of economics.

The people of Bali, has had advantages over human resources from abroad. This excess is the hospitality that has become nature. Even so, the user still needs to be equipped with support skills. The importance of improving the quality of human resources, so that people who are engaged in the economic field will always be criticized if it is judged not give satisfactory service.

This research attempts to develop instruments and models of employee performance evaluation graduate of Bachelor of Economics in Bali in support of sustainable development through an approach Bali human resources and wide range of indicators that influence, as the demands on the quality of performance which is needed to improve the quality of employees. Therefore, it is necessary to develop an instrument as well as a special evaluation model to determine the level of employee performance graduate of Bachelor of Economics, then the problem can be formulated as follows: (1) Whether the knowledge, attitudes in the workplace, job skills, motivation, and personality are indicators to predict employee performance graduate of Bachelor of Economics?; (2) How effective evaluation model for measuring employee performance graduate of Bachelor of Economics?; (3) How is the performance profile of graduates of Bachelor of Economics?.

The purpose of this study was to develop instruments and models of employee performance evaluation Economics graduates are most effective to be implemented in order to improve employee performance. The steps or stages are as follows: (1) Develop indicators used to predict this performance of employees graduate of Bachelor of Economics; (2) Develop an effective evaluation model for measuring employee performance graduate of Bachelor of Economics to be applied in order to improve the quality of services; (3) Determine the performance profile of Bachelor of Economics graduate employees in an effort to improve the quality of service.

II. THEORITICAL REVIEW

The new paradigm of higher education, known as tetrahedron shapes include economics, quality, accountability, accreditation, and evaluation. The curriculum is a key strategy in the empowerment of institutions of higher education. Curriculum development is based on the ideas, principles, and values that belongs institution. These values are usually set out in the statute. For long-term goals, then the educational institutions set the vision and mission that will be taken.

For the face of international competition, curriculum design should have a charge/content of international and national/local. International content is the key to international market entry, while the charge of national/local apply to enter the labor market in accordance with applicable skills qualification standards in Indonesia. Charge their national/local, in order to prepare graduates to compete in the national labor market.

Quality of Work Life (KKK) implies that is not the same for different people. For a Bachelor of Economics graduate employees, it can mean a reasonable level of wages, safe working conditions, and a supervisor who treat it in accordance with the status and dignity as human beings. For someone who just graduated from college it can mean an opportunity to move forward, creative tasks. And career success. According Suttle (Dessler, 1997) it means, "the circumstances in which a member of a labor organization is able to meet the needs of an important personal through their experience in the organization".

In short, the quality of working life means a state where employees can meet their needs of an important to work within the organization, and the ability to do so depends on whether there are any: (1) The treatment is fair, equitable, and supportive of employees; (2) The opportunity for every employee to use the full capabilities and opportunities to realize themselves, which is o be the person they feel can make it happen; (3) Open communication and mutual trust among all employees; (4) The opportunity for all employees to actively participate in making important decisions that involve their jobs; (5) Adequate and fair compensation; (6) A safe and healthy.

Brinkerhoff, et al (1986) explains that the evaluation is a process that determines the extent of educational goals can be achieved. According to Brinkerhoff , et al (1986) in implementing the seven elements of evaluation should be done: (1) focus on what will be evaluated (focusing the evaluation); (2) design/evaluation design (designing the evaluation); (3) collecting information; (4) analysis and interpretation of information (analyzing and interpreting); (5) create a report (reporting information); (6) setting/management evaluation (managing evaluation); and (7) evaluation for evaluation (evaluating evaluation)

The evaluation of the performance of employees graduate of Bachelor of Economics is intended to look back if the employee's performance has been in accordance with planning expected. The number of criteria that have been determined. The results of this evaluation can be used to determine the value or level of success of the program in terms of its effectiveness and efficiency, it can then take a decision whether the performance needs to be improved.

Evaluation model is the design model evaluations made by experts or experts evaluation (Tayibnapis, 2000). Draft evaluation model is used to evaluate the program. Many models can be used to evaluate a program. Philips (1991) grouped in seven types of evaluation model, namely: (1) The model of Kirkpatrick; (2) Model of The Bell System; (3) The CIRO models; (4) Models Saratoga Institute; (5) The models of IBM; (6) Models of Xerox; and 97) The model CIPP. Selection is determined by the evaluation model developed objectives and questions in the evaluation concerned.

This study uses the evaluation of the development of Kirkpatrick's model, which focuses on an assessment of employee performance (outcomes) and not the evaluation of the training program (process), assuming that employees are trained, and analyzes related components from various aspects (knowledge, attitudes in the workplace, job skills, motivation, personality, organizational satisfaction, job satisfaction, and customer satisfaction), to detect and give consideration to the repair of components or aspects that need improvement in achieving its objectives.

RAND Corporation on the development of this research is in the development stage research instruments in addition to using the Delphi method also using Focus Group Discussion (FGD), Delphi process focuses on the collection of a number of opinions of some individuals who are experts on the topics discussed. Delphi technique originally introduced by the RAND Corporation that aims to obtain a proper consensus among the experts on important issues without discussion face-to-face (Fazio, 1987). The use of FGD method in this study as a continuation of the Delphi method called earlier. Where FGD is informal, a discussion that was designed in small groups to obtain more in-dept qualitative information (Dean, 1994).

III. METHOD

This study is a research development and research evaluation is *ex post facto* as data collection. The main objective of the research is not to formulate development or test the theory, but developing effective results to be used in schools or other institutions. The purpose of the evaluation study is to facilitate decision-making with respect to the advantages/relative weakness of an action (Sumanto, 1990).

This study tried to determine a cause of what has already happened. Variable treatment on the design of this study is an occurrence that has happened. Therefore already occurred, the design of this study no treatment conducted by researchers. Why is the variable that has occurred because: (1) it is the variable can not be manipulated for reasons of ethics so that the objects sought studies that have experienced, or; (2) have occurred but have not been studied (Kountur, 2003).

The study design was based on System Performance Evaluation Model. This model starts from the view that the success of a system of employee performance is affected by many factors and the individual characteristics of the surrounding environment, the purpose of the system itself. Evaluation according to this model is intended to compare the performance of various dimensions of the system that is being developed by a number of specific criteria, to finally arrive at a judgment about the system description and are assessed.

These research subjects are Economics graduates who work in private companies/government agencies, state/enterprises, TNI/Polri in Bali some 1,000 people. Determining the size of the subject of research by proportionate stratified random sampling method. This technique is used when the subject of research has members/components are not homogeneous and stratified proportional (Sugiyono, 2004).

Fraenkel (1993) writes that for a descriptive study of the minimum required sample size of 100, correlation studies required at least 50 samples, whereas the experimental or causal comparative study requires a minimum of 30 samples. The sampling technique based on the formula Krejcie-Morgan where: $N = 1,000$; $\alpha = 0.05$; $\rho = 0.50$; $\chi^2 = 3.841$ obtained the sample size was 280 respondents (Isaac & Michael, 1982).

The trial of 260 respondents instrument intended to obtain the right instrument to be used in research, for potential respondents were absolutely right in research, to undertakings attempt repairs to the instrument that is less precise, and to determine the validity and reliability of the instrument, so it can be used to obtain the data required and can be accounted for.

Before statistical analysis applied to the underlying assumptions need to be proved first, in this study using a test for normality and linearity test.

Factor analysis was used to answer hypothetical number 1, is so to determine whether the indicators of knowledge, attitudes in the workplace, job skills, motivation, and personality can be used to measure employee performance graduate of Bachelor of Economics.

Structural equation modeling (LISREL) is one of the analytical technique in non-experimental studies were used to determine the direct and indirect influence of variables because of the causal variables (Pedhazur, 1982). LISREL used to answer hypoyhetical number 2, which maches the model to the performance evaluation of employees graduate of Bachelor of Economics.

To determine the profile of the employee's performance graduate of Bachelor of Economics also to determine the satisfaction level of the organization, the level of employee satisfaction and customer satisfaction levels using the analysis of the profile of Keeves. According Keeves (Hastuti, 2004), "the comparison between profiles of persons or groups on the same set of variables is known by profiles analysis". This means that the comparison between the profiles of individuals or groups of individuals on the same set of variables called profile analysis. Profile analysis is often used in the practice of research in the field of education such as a chart plotting, scores of individuals or groups in an assay both raw score and standard score. Descriptive statistics were used to answer the research question number 3, which is to know the profile of the employee's performance graduate of Bachelor of Economics.

IV. RESULTS AND DISCUSSION

Once past the trial stage, it can be concluded that the instrument measuring employees performance graduate of Bachelor of Economics has an index of validity and reliability that meets the requirements. The instrument is also quite easy to understand by the respondent.

A. Formulation of research hypothesis number 1: the indicators used to measure employee performance graduate of Bachelor of Economics consists of factors knowledge, attitudes in the workplace, job skills, motivation and personality.

Factor analysis was used to answer the research question. The measurement results KMO = 0.58 (> 0.5) and Bartlett test = 0.000 ($p = < 0.05$) means that the factor analysis can proceed.

Furthermore, the results of the factor analysis off all the factors that make up the latent exogenous variable is valid/invalid (> 0.30). This means that the indicators used to measure employee performance graduate of Bachelor of Economics consists of factors knowledge, attitudes in the workplace, job skills, motivation and personality.

B. Formulation of research hypothesis number 2: models of employee performance evaluation Economics graduates suitable to be applied on an industrial workplace.

LISREL analysis used to answer research questions. To determine the model fit to the data, use the following conditions: (a) A probability value > 0.05 ; (b) Goodness of Fit Statistics (GFI) between 0-1. Closer to 1 the better; (c) The value Root Mean square Residual (RMR), which is the smaller the better.

LISREL analysis is used to match the model with field data obtained. In addition, the LISREL analysis is intended as well to get the amount of employee performance parameter estimates (aspects of the service), which is used to supplement the information to answer the research hypothesis. LISREL analysis results show that: (a) Chi-square = 28.50; $df = 19$; and p -value of $0.743 > 0.05$; (b) Goodness of Fit Statistics (GFI) of 0.97420 approaching 1; (c) Square Root of the average Residual (RMR) = 0.044398 lies between 0-1. The results complied with. Thus, can be stated that the model fits the data, such as appears in Figure 1.

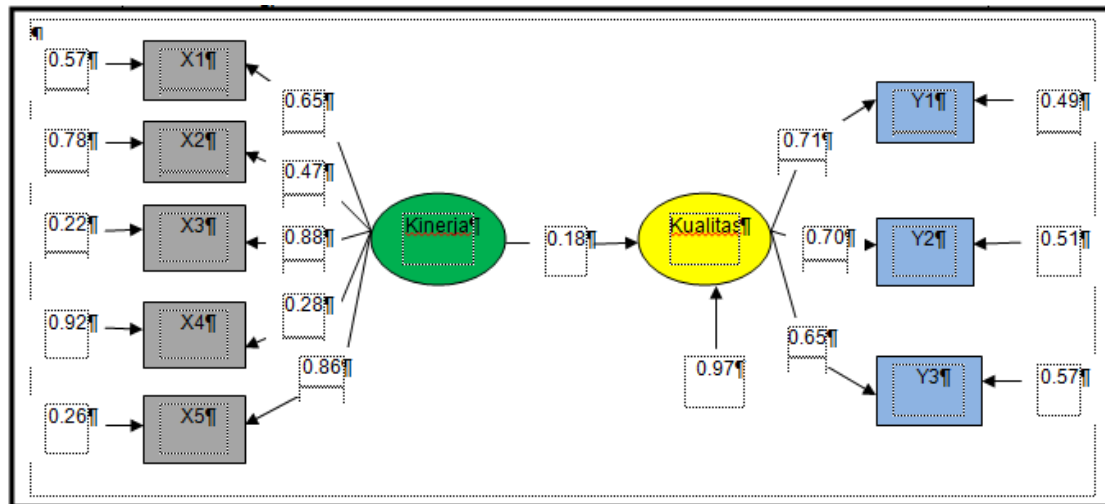


Figure 1. Employee Performance Evaluation Model Parameter Graduate Degree in Economics

C. Formulation study number 3: Economic performance profile of graduates working in the industry is good.

Descriptive statistics by SPSS 21.0 is used to answer the research questions. Output data analysis consisted of employee performance profile that is based on education and occupations.

V. CONCLUSIONS AND RECOMMENDATIONS

Research employee performance evaluation model graduate of Bachelor of Economics, carried out starting from the formulation and development of employee performance standards and its indicators, the development and implementation of performance measurement instrument to determine the structural equation models and determinants of employee performance profile graduate of Bachelor of Economics. Based on the results of our study, it can be deduced as follows:

1. Instruments employee performance graduate of Bachelor of Economics consists of five dimensions of performance, the dimensions of knowledge, attitudes in the workplace, job skills, motivation and personality.
2. The instrument consists of three service quality dimensions, i.e. the dimensions of organizational satisfaction, employee satisfaction, and customer satisfaction.
3. Model employee performance evaluation graduate of Bachelor of Economics developed in this study is the description of the characteristics of professional manpower in the industrial workplace.
4. The instrument has a degree of validity and reliability is quite high.
5. Descriptive analysis of the variables of knowledge, attitudes in the workplace, job skills, motivation, personality, organizational satisfaction, employee satisfaction, customer satisfaction in the high category.

VI. SUGGESTION

Based on the overall discussion and conclusions may be submitted suggestions to various parties follows:

1. To the Provincial Department of Education Bali.
Bali Provincial Department of Education should follow up on the development of performance standards of employees graduate of Bachelor of Economics is to formulate the minimum performance standards of employees and standardization, so that it can be applied to industrial work in Bali.
2. To the Bali Provincial Department Labor.
Department of Labor in employee performance improvement economics graduates, should always conduct coordination and cooperation with the Bali Provincial Department of Education, so that the achievement of a Bachelor of Economics graduate employee performance can be optimized.

3. In order to achieve higher performance, presumably a very important consideration is the internal and external factors which consist of: career commitment, work experience, leadership and work culture because it will increase the confidence of employees in improving their performance against the company, so the goal company to create the best services can be realized.
4. Economics is a science that should be “**professional**”, can stand alone and apart from the other sciences in an effort to improve productivity and performance of graduates of Bachelor of Economics.

REFERENCES

- [1] D.L. Dean, How to use focus groups. Dalam S.W. Joseph, H.P. Harry, & N.E. Kathryn (Eds), Handbook of practical program evaluation (pp. 338-349). San Fransisco: Jossey Bass Publishers, 1994.
- [2] E.J. Pedhazur, Multiple regression in behavioral research, New York: Holt, Rinehart and Wiston, Inc., 1982.
- [3] G. Dessler, Manajemen personalia. (Terjemahan D. Agus). Florida: Reston Publishing Company, Inc, 1997. (Buku aali diterbitkan tahun 1984).
- [4] H.S. Budi, Pengembangan instrument kompetensi tenaga lapangan Dikmas (TLD) di Provinsi Jawa Tengah, Tesis magister, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta, 2004.
- [5] J.J. Phillips, Handbook of training evaluation and measurement methods, Houston: Gulf Publishing Company, 1991.
- [6] J.R. Fraenkel., & N.E. Wallen, How to design and evaluate research in education, New York: McGraw Hill Inc., 1993.
- [7] K. Ronny, Metode penelitian, Jakarta: PPM, 2003.
- [8] L.S. Fazio, The Delphi: Education and assessment in institutional goal setting. Dalam S.C. David & L.W. Mark (Eds.), Evaluation studies: Review annual, 11 (pp. 287-297), California: Sage Publications, Inc., 1987).
- [9] R.O. Brinkerhoff., D.M. Brethower., T. Hluchyj., et al, Program evaluation a practitioner's guide for trainers and educators, Western Michigan: Kluwer-Nijhoff Publishing. 1986.
- [10] S. Isaac., & W.B. Michael, Handbook in research and evaluation: For education and the behavioral sciences (2nd ed), EdITS Publishers: California, 1982.
- [11] Sugiyono, Metode penelitian bisnis, Bandung: CV. Alfabeta, 2004.
- [12] Sumanto, Metodologi penelitian social dan pendidikan, Yogyakarta: ANDI OFFSET, 1990.
- [13] T.Y. Farida, Evaluasi program, Jakarta: PT. Rineka Cipta, 2000.

HYPOTHETICAL MODEL DEVELOPMENT OF ELECTRICAL TORSO LEARNING MEDIA CIRCULATION SYSTEM FOR STUDENTS SKILL FORMATION OF CRITICAL THINKING AND SCIENTIFIC ATTITUDE SENIOR HIGH SCHOOL IN LAMPUNG TIMUR

Ririn Noviyanti, Sisca Puspita Sari Nasution
Mahasiswa Magister Teknologi Pendidikan Universitas Lampung
Mahasiswa Magister Teknologi Pendidikan Universitas Lampung
ririnnoviyanti943@yahoo.com

Abstract - This research aimed to develop hypothetical model development of electrical torso learning media circulation system to form critical thinking skill and scientific attitude. The methodology of this research were conducted by theoretical and empirical studies, through the planning stages of development with literature studies and field studies, and hypothetical model development stage. The result of this research was about hypothetical model containing stage planning with field studies and literature studies based on the study of theoretical and the result of relevant research, then stage of development hypothetical model as arranging the draft by identification of students requirements, define students requirements priority, designed learning programme such as syllabus and lesson plan, define the topic to arrange lesson media development, and define kind of lesson media, then do the draft test of lesson media through expert validation from material side and lesson media and then the revision draft of lesson media development from feasibility test of lesson media

Keywords: *Learning Media Development, Critical Thinking Skills, Scientific Attitude*

I. PREFACE

Achievement of competence is closely related to the learning process that takes place in the classroom, because the learning process can influence the evaluation of learning outcomes. One factor that can support the learning process is to create an interesting and fun class atmosphere so that it can build students interest to participate actively in learning. To create an interesting class atmosphere the teacher can construct learning strategies by utilizing various media as a source of learning (Wina Sanjaya, 2013).

The learning process in the classroom will be more meaningful if students can personally experience things they learned (Aqib, 2014). In addition, the learning process also has a goal to lead the students to apply their knowledge in daily life. The concept of learning that associate the teaching materials with direct experience in the real life is needed in order to help students construct knowledge. For some of the subject which has complex material and tend to be difficult to understand indispensable alternative learning resources. One of them is the media that according to the conditions and needs of learners, the use of instructional media in the learning process can arouse desire and new interest and stimulation even bring psychological influences to the students.

It's relevant with the concept of contextual learning that focuses on the relevance of learning activities and their relation to everyday life (Hutagaol, 2013). According to Andi and Sari (2016) creativity of teachers in developing of learning media design as a source of learning is necessary to create learning that is not passive, but active. Teachers should have the ability to utilize a

variety of learning resources which available at the school, even though its simple but able to support the achievement of learning objectives. With creativity, teachers can develop an instructional media design in accordance with the development of science and technology. This is reinforced by the opinion of Arsyad (2015) that the media is an integral part of the learning process in order to achieve the goals of education in general and the purpose of teaching in schools in particular.

According to Iskandar (2016) science learning in general aims to develop curiosity, positive attitude, and awareness of the existence of mutual influence relations between science, environment, technology, and society as well as conduct scientific inquiry to cultivate thinking ability, behave and scientific attitude and communication. The purpose can be acquired through a learning process that can develop critical thinking skills and scientific attitude. This is reinforced by the opinion of Hidayati (2010) that by developing thinking skills explicitly and combine them with learning materials (curriculum) can help students to become critical and creative thinkers effectively.

Learning that can improve critical thinking skills and scientific attitude can actually be done in the classroom, but because of the limitations that these two aspects often go unnoticed, especially biology learning is closely associated with the various processes of the body that are difficult to observe, for example the material of circulation system. This limitation can be obstacles and causing the learning process becomes not maximal so that its necessary development of props to visualize the material which is tends to abstract while improving critical thinking skills and scientific attitude of students. This is reinforced by the opinions of Daryanto (2013) that the media serves to obtain a clear image of objects / things which difficult to observed directly because it does not allow to be observed directly. The use of three-dimensional learning media such as Electric Torso designed as props to provide hands on experience to students (Tapestry, Supriyanto, Marianti: 2012).

Based on observations in some SHS in East Lampung availability of props as a learning media is still very minimal, especially props for the subjects biology to develop critical thinking skills and scientific attitude of students. This resulted in students having difficulty to understand the material. Therefore it is necessary to develop a props of three-dimensional learning media, in this case a Electric Torsoto visualize the blood circulation in humans. The use of three-dimensional media will provide hands on experience to students, presenting material in concrete way and avoid verbalism (Moedjiono, 1992).

From the research of Permadani, Supriyanto, Marianti (2012) concluded that the Electric Torso circulatory system developed effective learning. Therefore, the authors are interested in doing a research on the Hypothetical Model Development of Electric Torso Circulatory System High School in East Lampung.

II. RESEARCH METHOD

This study was developed from the results of several theoretical review and relevant research studies. The method of this research is adopted from Affiyani research methods (2013) modified into, (1) preparation of hypothetical model development, by reviewing the literature related to the theory, concepts, and the results of relevant studies to support a preface study of the field, further empirical studies, carried out by field studies through observation and interviews related to the utilization of lesson media as a source of learning.

According to Siti Hamidah (2013) from literature and field study obtained the concept of theoretical and empirical data. The theoretical concepts will direct the formation of a hypothetical model lesson media development. The next stage, according Affiyani (2013), (2) Designing a hypothetical model development of electrical torso.lesson media. The development phase is done by processing the information obtained in the preface study. The activities pursued in the development stage are reviewing a comprehensive analytical and critical to all information and data. Analytical critical studies done by discussion and brainstorming. Whole discussion of activities directed at the establishment of the learning model which is based on a new paradigm of education in Indonesia.

In this stage continued with hypothetical model development. According to Muhammad Taufik, et al (2010) after having obtained a hypothetical model development. The next step is the completion model stage, then generates the final model, and evaluation.

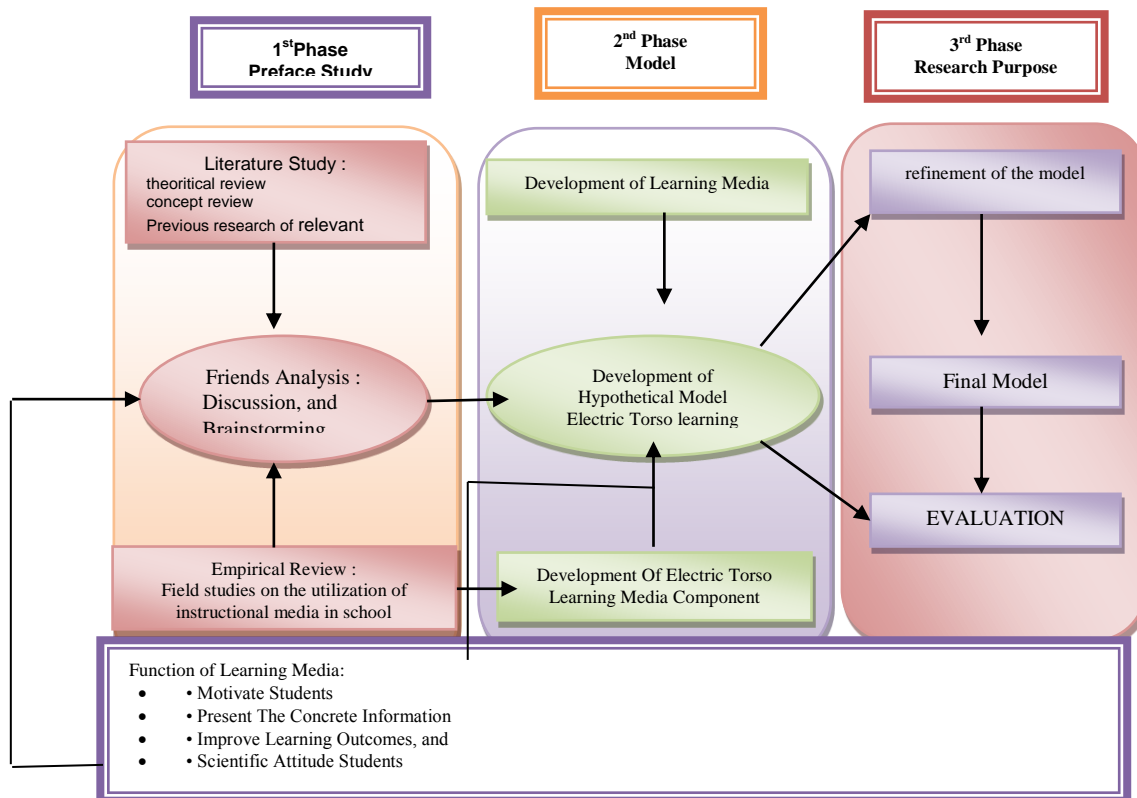


Figure 1. Kerangka Berpikir Model Penelitian Hipotetik

III. RESULT AND DISCUSSION

The research of Learning Media Development Elektrik Torso Circulation System, conducted through in three stages, namely:

A. Assessment of Learning Media Factual Conditions of Use (Existing) And Literature (Pre Existing Assessment)

Assessment of the utilization of factual learning media focused on the type of media that used by teachers, and utilization. Based on interviews obtained the following information:

1. Types of learning media used by teachers

Types of learning media that used by the teacher is still very limited such as, text books, worksheets, picture, and ourselves as a model. The availability of learning media is still very minimal, especially props for the biology subjects to develop critical thinking skills and scientific attitude of students. This resulted in students having difficulty to understand the material. Moreover, the material cannot be observed directly

2. Theoretical study

Theoretical studies or literature study is a preliminary descriptive study, as the basic construction of hypothetical model. Description of the results of theoretical studies were developed as behaviorism learning theory, Vigotsky cognitive social learning theory, information processing, Gagne's learning theory, Ausubel's learning theory, Bruner's learning theory, constructivistic learning theory, humanistic learning theory.

According Utami, Purnomo, and Rizal (2016) learn according to behavioristic theory is a change in behavior as a result of interaction between stimulus and response. In the other words, learning is the change experienced by students in terms of its ability to behave in a new way as a result of interaction between stimulus and response (Suryadi, 2016)). A person is considered to have learned something if he can show change in his behavior (Grace, Smith, and Rahim, 2016). According to this theory, the most important thing is input in the form of stimulus and output or output in the form of response. Stimulus is what the teacher give to the student (Umar, and Sulandjari, 2016) for example multiplication tables, pictures, props, guidelines, or way to help

students learn. While the response is a reaction of students to the stimulus provided by the teacher (Budiningih, 2005).

According to Vygotsky (Baharuddin and Wahyuni, 2007) that learning begins when a child in the development of proximal zone, which is a level achieved by a child when doing social behavior. This zone can also be defined as a child who can not do anything alone but requires the help of a group or an adult. In the study, proximal zone can be understood as the difference between what can be done by someone with the group or with the help of an adult.

In Vygotsky's theory is explained there is a direct relationship between cognitive socio-cultural. The quality of student thinking is built in the classroom, while the social activities developed in cooperation between students with other students who are better under the guidance of teachers (Isjoni, 2010).

According to the theory of information processing, learning is seen as an information management processes in the human brain that stated by Gagne. According to this theory, learning is seen as a process of information processing in the human brain. According to the Ausubel's learning theory, students will learn well if the the content of previous lessons well defined and presented to students (advance organizers). Meanwhile, Bruner proposed a theory called free discovery learning. This theory explains that the learning process will run well and creative if the teacher gives students the chance to find a rule (concepts, theories, definitions, etc) through examples that illustrate the rule that became a source of students.

Constructivist learning theory to understand learning as a process of formation / construct knowledge by the learning itself (Wahyuni et al, 2016). In constructivist flow, knowledge is understood as a continuous formation by a person who at any time was reorganized for their new understanding (Febrianti, 2016). Knowledge is not the ability of the facts of a reality that is being studied, but as a cognitive construction of the object, experience, or environment (Siregar, 2016).

Meanwhile, in the view of the humanistic learning theory, learning process is done by providing the freedom for the individual. The learner is expected to take its own decisions and take responsibility for the decision (Gredler, 2013).

Based on the constructivist paradigm, the media occupies a strategic position in order to realize optimal learning event (Daryanto). Optimal learning event is one of the indicators to achieve optimal learning results anyway. Good learning outcomes / optimal is also a reflection of quality education. Quality education requires teachers resources who are able and ready to contribute professionally (Heinich et.al., 2002; Ibrahim, 1997; Abraham et al, 2001).

Professional teachers demanded to be able to select and use different types of media that exist around us, or even able to create or develop an appropriate instructional media with the purpose of learning.

B. Develop a prototype of hypothetical model development of electric torso learning media circulation system

Based on the theoretical review, then obtaining prototype of Hypothetical Model Learning Media Development Electrical Torso, namely:

1. Plan

Planning Activity of development learning media electric torso includes:

a. Preface Study include literature studies and field studies.

In a literature study researchers examined the theories or concepts related to the type of media that is used by teachers, as well as previous relevant research findings. Field studies done to find the type of media that is suitable to be developed, identifying the real conditions supporting or inhibiting factor (condition of capability, performance of teachers and students, as well as the quantity and quality of the facilities and infrastructure)

2. Hypothetical model development of electric torso learning media circulation system

a. Preparation of learning media draft development

In the drafting of instructional media development, thing to do are, identify student's learning needs, and then determine the priority learning needs, designing learning programs (syllabi and lesson plans), the determination of topics for the preparation and development of instructional

media, the determination of the type or class of instructional media, organizing content / materials and materials required (specification of the content / materials and materials required).

b. Testing learning media draft development

Preparation of a draft instrument testing through the development of instructional media content expert validation/materials and media. The next is feasibility test due of learning media draft. There are five aspects to be measured for measuring the development of learning media: feasibility component aspects of the content, presentation aspects, the quality aspects of learning media, aspects of critical thinking skills and scientific aspects of the student's attitudes. Respondents are given the form of a scale of 1 to 5 with the following conditions: STS = Strongly Disagree (score = 1), TS = Disagree (Score = 2), N = Neutral / Common (Score = 3), S = Agree (score = 4), SS = Strongly Agree (score = 5)

1) Aspect of the feasibility component contents

Aspects of the feasibility component content, is one important element in the development of instructional media. Good learning media is media that is created in accordance with the purpose of learning. There are four indicators used to measure the feasibility of the component aspects of the content. As shown in Table 1

Table 1. Indicators of component aspects of the contents feasibility

No	Aspects	statement	Score
1	Depth Content	The scope of materials relevant to education, not only the material contained in the book of textbooks, materials conformity with relevant concepts with SK/KD, the material can increase knowledge	
2	The relevance of the learning purpose with competency standard/basic competencies/curriculum	Delivered learning purpose that relevant to SK/KD/Curriculum	
3	Truth and accuracy concept	The concept of the material presented is true and correct	
4	Truth and correctness of the theory	The theory contained in the material presented is very true and correct	
Total			

2) Presentation of media aspect of learning

In the process of learning, the media has a function as carriers of information from the source to the receiver (Daryanto, 2013). Presentation of learning media is one aspect that is important in the process of the information carrier. There are five indicators used to measure aspects of the presentation of instructional media. As stated in Table 2.

Table 2. Indicators presentation aspects of learning media

No	Statement	Score
1	Encouraging students to know the contents of instructional media	
2	Stimulate the participation of students for independent study and group	
3	Presentation is communicative and interactive	
4	The material is described in a coherent, systematic, logical flow of clear instructional clear material flow	
5	Simulation experiments can be done clearly and independently	
Total		

3) Quality of the learning media

According Crititos 1996 (In Daryanto, 2013) is one of the media communication component, namely as a messenger from the communicator to the communicant. Messages can be conveyed through the media with both the quality pembelajarannya media should also be good. There are five aspects to measure the quality of media pembelajaran, as presented in Table 3.

Table 3. Indicators of quality aspects of learning media

No	Statement	Score
1	Encouraging students to know the contents of learning media	
2	Stimulate the participation of students for independent study and group	
3	Presentation is communicative and interactive	
4	The material is described in a coherent, systematic, logical flow of clear instructional clear material flow	
5	Simulation experiments can be done clearly and independently	
Total		

4) Aspects of Critical Thinking Skills

Critical thinking is an attitude to think deeply about issues and things that are within the reach of a person; knowledge of the methods of examination and logical reasoning; and a kind of a skill to apply these methods (Glaser, 1941). There are five indicators and 8 keterampilan aspect to measure critical thinking, as presented in Table 4.

Table 4. Indicators aspects of critical thinking skills

No	Indicators of critical thinking skills	Aspects of critical thinking skills	Score
1	Give a simple explanation	focusing questions	
		Analyze the question and ask	
		Answering a question about an explanation or question	
2	Build basic skills	Mempertimbangkan apakah sumber dapat dipercaya at Consider whether the source is reliable or not	
3	Conclude	Deducing ctivities and consider the results of deduction, induce and consider the results of induction	
		Create and deduce the value of the consideration	
4	Further explanation	Identify terms and d consideration efinitionsand also the dimensions, and identifying assumptions.	
5	Set the strategies and techniques	Determine the action and interaction with others	
Total			

5) Aspects of Scientific Attitude

Attitudes can be defined as the internal capabilities that play a role In taking action. Where actions to be selected, depending on his attitude towards gain or loss, good or bad, satisfactory or not, of an action does (Slameto, 2010).

The attitude has become one of the factors that affect learning outcomes. The attitude in learning Science is often associated with attitudes toward science, or better known by the scientific attitude. Aspects that are used to measure students' scientific attitude, as presented in Table 5.

Table 5. Indicators scientific attitude

No	Indicator	Aspect	Score
1	Want to attitude	Seeking answers from the teacher questions through other sources of information other than that provided by the teacher	
		Enthusiastic in participating in learning activities	
		The focus of the explanation given by the teacher	
2	Open minded attitude	Listen when friends present their views, ask and answer without being interrupted friend proposes to ask or answer questions.	
3	Impartiality	Did not cheat other people's work	
Total			

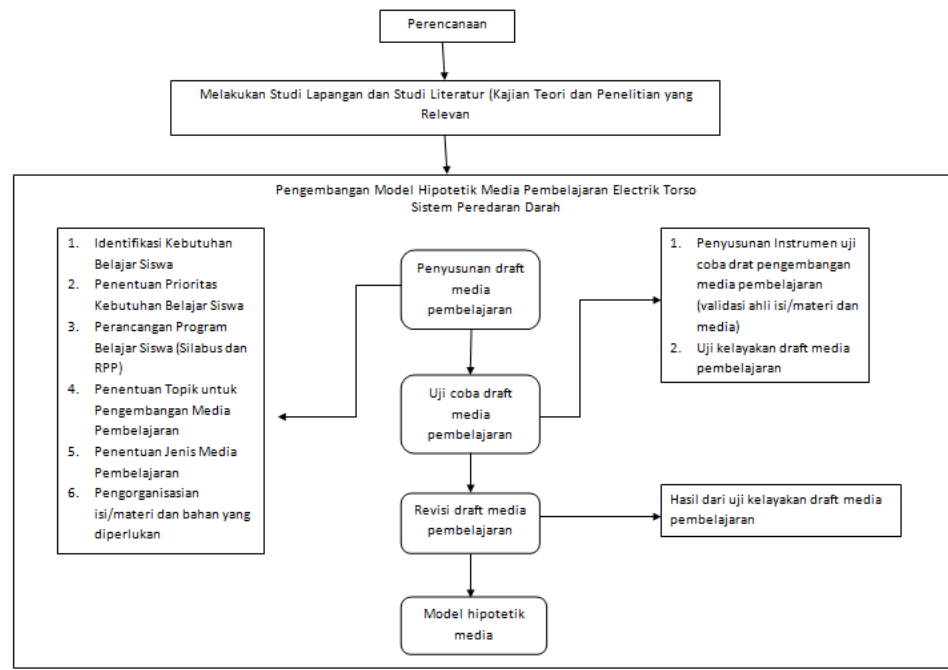
6) Revised Draft Media Development Learning

Input or data obtained from the results of due diligence draft learning media

7) Learning Media Development hypothetical model Electric Torso

Draft development of instructional media that have been revised to produce a prototype model of a hypothetical development of instructional media electric torso circulatory system.

Here is a hypothetical model of instructional media development electric torso circulatory system:



IV. CONCLUSION

Based on the results of the preliminary analysis it is found that there is no factual development of instructional media models elektrik torso circulatory system SMA in East Lampung regency. Based on the study of the learning technology and literature studies (field) then the resulting hypothetical model of instructional media development elektrik torso circulatory system.

REFERENCES

- [1] Adhi P, S., & Sari, L. P. (2016). Penerapan Model Pembelajaran Contextual Teaching and Learning (CtL) Berbantu Media Gambar terhadap hasil Belajar, Kinerja Guru Dan Aktivitas Siswa Ips Semester 2 Kelas Vi Sdnegeri 1 Bugo Jepara. *Media Penelitian Pendidikan*, 9(2 DESEMBER).
- [2] Aqib, Zainal. 2014. *Model-Model, Media, dan Strategi Pembelajaran Kontekstual*. Yrama Widya: Bandung
- [3] Arsyad, A. 2015. *Media Pembelajaran*. Raja Grafindo Persada: Jakarta
- [4] Baharuddin dan Wahyuni, N. (2007). *Teori Belajar dan Pembelajaran*. Ar-Ruzz Media Group: Yogyakarta
- [5] Budiningsih, Asri. 2005. *Belajar Dan Pembelajaran*. Rineka Cipta: Jakarta
- [6] Daryanto. 2013. *Media Pembelajaran Perannya Sangat Penting Dalam Mencapai Tujuan Pembelajaran*. Gava Media: Yogyakarta
- [7] Febrianti, W. A. (2016). *Penggunaan Model Pembelajaran Inkuiri Terbimbing Untuk Menumbuhkan Sikap Rasa Ingin Tahu Dan Percaya Diri Serta Meningkatkan Hasil Belajar Siswa* (Doctoral dissertation, FKIP UNPAS).
- [8] Glaser, E. 1941. *An Experience in the Development of Critical Thinking*. Advanced School of Education at Teacher's College: Columbia University
- [9] Gredler, M. E. (2013). *Learning and instruction: Teori dan aplikasi*
- [10] (edisi keenam). Kencana Prenada Media Group: Jakarta.
- [11] Hamidah, S. (2013). Model Pembelajaran Soft Skill Terintegrasi pada Siswa SMK Program Studi Keahlian Tata Boga. *Jurnal Pendidikan Vokasi*, 2(1).
- [12] Heinich, Molenda, Russel, Smaldino (1996). *Instructional Media and Technologies for Learning*. Prentice Hall, Englewood, New Jersey
- [13] Hutagaol, K. (2013). Pembelajaran Kontekstual untuk Meningkatkan Kemampuan Representasi Matematis Siswa Sekolah Menengah Pertama. *Jurnal Infinity*, 2(1), 85-99.
- [14] Isjoni. 2010. *Cooperative Learning*. Alfabeta: Bandung.
- [15] Iskandar, S. M. (2016). Pendekatan Keterampilan Metakognitif Dalam Pembelajaran Sains Di Kelas. *ERUDIO (Journal of Educational Innovation)*, 2(2).

- [16] Muhtadin, M. (2016). Pengaruh Pelaksanaan Praktik Secara Terintegrasi Terhadap Prestasi Belajar Dasar–Dasar Otomotif Siswa Kelas X Jurusan Teknik Kendaraan Ringan Smk Piri 1 Yogyakarta Tahun Ajaran 2010/2011 (Doctoral dissertation, UNY).
- [17] Moedjiono. 1992. *Media Pendidikan III : Cara Pembukaan Media Pendidikan*. P3G. Departemen Pendidikan Nasional: Jakarta.
- [18] Permadani, K. G., Supriyanto, S., & Marianti, A. (2012). Pengembangan Electric Torso Pada Pembelajaran Sistem Peredaran Darah untuk SMP. *Unnes Journal of Biology Education*, 1(3).
- [19] Pramono, A. (2013). Pengembangan Model Bimbingan Kelompok Melalui Teknik Psikodrama untuk Mengembangkan Konsep diri Positif. *Jurnal Bimbingan Konseling*, 2(2).
- [20] Rahmat, A., Smith, M. B., & Rahim, M. (2016). Perilaku Hidup Sehat Dan Prestasi Belajar Siswa Sekolah Dasar. *Psychopathic*, 2(2), 113-122.
- [21] Sanjaya, W. (2013). *Strategi Pembelajaran: Berorientasi Standar Proses Pendidikan*. Kencana Prenadamedia Group: Jakarta.
- [22] Siregar, J. (2016). Penerapan Metode Simulasi Untuk Meningkatkan Aktivitas Dan Hasil Belajar Pkn Siswa Di Sd Negeri 187/Iv Kota Jambi.
- [23] Slameto. (2010). *Belajar Dan Faktor – Faktor Yang Mempengaruhinya*. Rineka Cipta: Jakarta
- [24] Suryadi, S. (2016). Peranan Perkembangan Teknologi Informasi Dan Komunikasi Dalam Kegiatan Pembelajaran Dan Perkembangan Dunia Pendidikan. *Informatika*, 3(3), 63-75.
- [25] Taufik, M., Sukmadinata, N.S., Abdullah, I., Tumbelaka, B. Y. (2010) Desain Pembelajaran Untuk Meningkatkan Kemampuan Pemecahan Masalah Dalam Pembelajaran IPA (Fisika). *Jurnal Berkala Fisika*, 13(2.)
- [26] Umar, H. R., Sulandjari, S., & Si, M. (2016). Pengaruh Penerapan Model Pembelajaran Problem Based Learning Pada Materi Bumbu Dasar dan Turunannya dalam Makanan Indonesia Terhadap Hasil Belajar Siswa SMK Negeri 2 Mojokerto. *Jurnal Mahasiswa Teknologi Pendidikan*, 5(1).
- [27] Utami, A. C., Purnomo, E., & Rizal, Y. (2016). Efektivitas Vct Dan Ctl Dalam Menumbuh Kembangkan Sikap Terhadap Wirausaha Siswa. *Jee (Jurnal Edukasi Ekobis)*, 4(4).
- [28] Wahyuni, S., Darsono, D., & Pargito, P. (2016). Pengembangan Model Pembelajaran Inquiry Untuk Meningkatkan Kemampuan Memecahkan Masalah Sosial Di Masyarakat. *Jurnal Studi Sosial*, 4(1).

DEVELOPING A CREATIVE THINKING ASSESSMENT MODEL FOR KINDERGARTEN TEACHERS

Risky Setiawan

IKIP Veteran of Semarang Indonesia
setiawan.risky@gmail.com

Abstract - This research and development aims (1) to develop an assesment instrument for creative thinking, and (2) to find out the characteristics of the creative thinking assessment instrument for the kindergarten teachers in the province of Central Java. The study used the Research and Development approach with ADDIE instructional model (Analyze-Design-Develop-Implementation-Evaluation) by Melinda. The subjects of the research were kindergarten teachers in Central Java selected by using purposive random sampling consisting of 281 kindergarten teachers. The product development result had been tested specifically to the respondent and then be tested generally to 281 respondent. The construct validity analysis result applied Confirmatory Factor Analysis (CFA) showed 39 valid item, and reliability coefficient Alpha Cronbach and composite score shows good categorized (>0.7). The results of the research show as the following: 1) the kindergarten teachers creative thinking constructive assessment instrument covered 3 components (teacher's professionalism, convergent thinking, and divergent thinking), 2) the characteristic development product is; a) on each instrument: (1) professionalism teacher instrument: (a) unidimensional, (b) suitable with the model GPCM (Generalized Partial Credit Model), (c) the optimal estimation TIF on theta -2.02; (2) divergent thinking instrument: (a) unidimensional, (b) suitable with the model GPCM, (c) the optimal estimation TIF on theta 0.25; (3) convergent thinking instrument: (a) unidimensional, (b) suitable with the model 2PL and 3PL, (c) the optimal estimation TIF on theta 1.5; b) the developed model was identical to the theory (H_0 is rejected), c) the developed product's effectiveness is categorised as high.

Keywords: *assessment, the IRT, divergent thinking, convergent thinking, kindergarten teacher professionalism.*

I. INTRODUCTION

Early childhood intelligence potential can be maximized through effective and fun learning. The Potential of Early childhood intelligence consists of: (1) creative and emotional intelligence, (2) physiology intelligence, and (3) mathematical and language intelligence. Thus we need some effective measures to optimize the three of Early childhood intelligences. The application to increase early childhood intelligence can not be separated from the role of an educator, especially teachers. Kindergarten teacher is taking a great control over the potential to be developed for Early childhood. The role of the teacher as the closest to the children in schools should be able to create a pleasant atmosphere and become a person whose liked by the children. The role of the teacher as a friend, models, motivators and facilitators will make children enjoy coming to school and make the learning process so meaningful. Therefore it's demanded maturity that require willingnes and ability, both intellectually and in excellent condition. Professionalization like this should be viewed as a process that is continously.

The relationship between level of education and creativity are very closely. Torrence theory says that one of the indicators that affect the development or expression of creative thinking is a person's education level [1]. The professionalism of a teacher is not impossible, despite the survey data of the political and econimic risk country (PERC), a consultancy in Singapore in 2001 ranked Indonesia 12th of 12 countries in Asia in terms of the teachers quality. While it was said by decree number 19 year 2005 declared that a kindergarten teacher should have four basic competencies (professional, pedagogical, personality, and social) and additional competence,

namely: 1) creativity and the sense of art as an adjustment between learning and the child's needs, 2) understanding of development theory and its implications [2]. In addition, the kindergarten teacher also has additional obligations include: 1) be an example for the character building, 2) develop learning plans according to the development stage, 3) managing the play activities according to the children development and interests, and 4) carry out the assessment in accordance with the children capabilities achieved [3].

According to Dweck and Leggett in Angela Anning et al [4], it is very important in the development of mastery learning in early childhood. Many facts prove that the professional and experienced teachers in the world of education will provide better response and enhance the learning effectiveness in the classroom. Learning effectiveness is still difficult optimized since the kindergarten teachers' salaries is low. The Chairman of Commission IV DPRD Purbalingga assess welfare honor given to kindergarten and Madrasah Diniyah (Madin) teachers in Purbalingga is inhumane. Currently they are only paid Rp 75.000 for kindergarten teachers and Rp 70.000 for Madin teachers [5]. The poor quality assumptions of the kindergarten teachers is justified by Syahwal Gultom, BPSDMPK Head (Human Resources Development Agency of Cultural Education (Badan Pengembangan Sumber Daya Manusia Pendidikan Kebudayaan), Kemdikbud. He said, "The competency tests results conducted over the last three years shows the quality of kindergarten teachers in Indonesia is still very low" [6].

Haryati [7] states that the assessment system used to improve the quality of education is continuous assessment. Reality assessment of children now only focus on creativity as a factor of the power amplifier child cognition. Moreover, the power of affection should be preferred because the child patterned attitude will be faster impact on the absorption of knowledge-the basic knowledge children. Glazer, 2009 [8] argues that knowledge and art is a fusion of a creative person. The creativity of teachers in kindergarten regarded as a potential environmental and capital compared to the second aspect of the professionalism of an educator. Theory of Belkhasdas [9] says: "Creative teaching to Increase students' learning and achievement", a creative learning will improve learning and knowledge learners. Or assumed learning undertaken by a kindergarten teacher becomes the main factor of increasing knowledge of learners.

The importance of creativity kindergarten teachers need to be proven with a device that can measure precisely. The test is one way to measure the results and effectiveness of learning. Effectiveness and learning outcomes in the classroom will determine the quality of education in general. Teacher competency tests only give an overview of cognitive ability of teachers. According Plucker in Stenberg [10], divergent thinking test is a method often used to measure a person's creative berpikir process. A number of creative tests have been developed and implemented, such as tests of Torrence to measure creative thinking that has the form of verbal and figural forms [11]. Theory Guilford in Bloomberg, [1] confirms factor in "Structure of Intellect" classify the person's intellectual operation can be seen from the way of thinking Convergent and Divergennya. From the description concluded that the measurement of the ability to think creatively a kindergarten teacher can be done by arranging two types of tests are tests of convergent and divergent test.

Implementation of measurement and assessment is important because the creativity of teachers in kindergarten element is the ability of the most influential children's creativity. Impact, a creative child can perform any activity with optimal play, especially on activities both in the classroom and outside the classroom. Therefore, experience and creativity is very important to measure. Measurement or measurement is one of the programs conducted to determine the standard of behavior or ability includes measures on a predetermined scale. From the description above explanation can be concluded measure creative thinking abilities of teachers is very important to do. Background Based on the description can be elaborated formulation of the problem, namely; 1) how to construct the assessment instrument kindergarten teachers creative thinking ?, and 2) how the assessment instrument product characteristics of creative thinking kindergarten teacher?

II. RESEARCH METHOD

The research method is a research and development by developing creative thinking abilities gauges kindergarten teacher in Indonesia . The model used in the development of ADDIE model is developed by Molenda 2003. ADDIE model is a model that adopts the IPO evaluation model (input, process, output), which consists of the step (1) Analyze, (2) design, (3) develop, (4) implementation, (5) evaluation [12]. The instrument further arranged in tested to make the contents and compiling of test and non-test proper and fit. Advanced analysis quantitatively

analyzed with Item Response Theory. IRT Analysis on Remote Asisted Test Instruments, Torrance Test of Creative Thingking (CCTC) and non-testing instruments Learning Case in the learning process by the method of Generalized Partial Credit Model (GPCM).

To determine the quality of test instruments to analyze both qualitative validation expert (expert judgment) in terms of content aspect, construction and language as well as quantitatively through the testing process. Limited testing done to see quality items using Quest program. The subject of limited testing will be done for 75 kindergarten teachers in the city of Semarang. Expanded trials conducted to prove the reliability of the instrument and the effectiveness of the instrument. By adding the subject will be able to increase the validity and reliability of the instrument. Determination of the number of samples using purposive random sampling technique that is determining the number of samples based on a certain amount. Total population of all kindergarten teachers and RA (Raudlotul Athfal) is as much as 48 987 teachers in 35 districts in Central Java (Central Java Statistical Information, 2015). According to the table Kracjie total population of 48 987, the obtained experimental data $n = 281$ respondents.

The method used to identify the problems and needs is a test, observation, distribution of questionnaires, documentation of research results, and discussion. In addition, data obtained through expert advice snapping-measurement, all TKan and prospective users, both in oral form (input in meetings) and a questionnaire (written input). The method further discussions are recorded and implemented in the assessment of creativity of teachers that have been refined to being transformed into a form of creative thinking assessment instrument is a kindergarten teacher.

The research instrument based on objective assessment of Developing Creative Thinking Kindergarten Teacher as follows.

Table 1. The Instrument of Creative Thinking

No	Variable	Categorical	Indicator
1	Personal Test	Pedagogic, Professionalism, Social and Personal	Learning Instructional
2	Divergent Test	Use and different Founded and combine Complete and conclude	Figural test Choose and pick theme
3	Convergent Test	Ask and guess Couse and consequences Unusual and suppose to problem solving	Verbal Test

III. RESULTS AND DISCUSSIONS

A. Assessment of Initial Product

1. Indicators Fit Assessment Model

The details of the output lisrel results are as follows: 1) Value P-Value = 0.067 (> 0.05), 2) Value RMSEA = 0.052 (< 0.08). 3) The value of GFI was 0.46, (a value between 0 and 1). It can be concluded instrument being tested is fit and match the model.

2. Value T-Value and Loading Factor

Significant votes whether the parameters or relationships between variables can be seen from the t-value. The significance of the parameters to see the relationship between variables can be seen in the regression equation in LISREL output with the following calculation formula.

$$t - value = \frac{\text{Regression Estimate}}{\text{Standar Error}} \dots\dots [13]$$

The estimation results of the T value is calculated through a formula by dividing the standard error of regression estimation of measurement if the value is greater than 1.2, the correlation between variables is good or significant. Whereas if the value is less than 1.2, the correlation between variables is less well or not significant. From the whole instrument contained 22 items that have a strong relationship between the variables that 37% had a t-value indicators are very good. The next step is to test the construct validity. Test the construct validity can be seen through convergent validity value (the value of the loading factor). Determination fit model aims to see whether or not fit items. Rated loading factor at the beginning of the study (research and development) is ≥ 0.50 (Gozali), while according to Fernandes, 1984, p.28, when the value of λ (Lamda) or loading factor greater than 0.3, the point of such instruments considered valid.

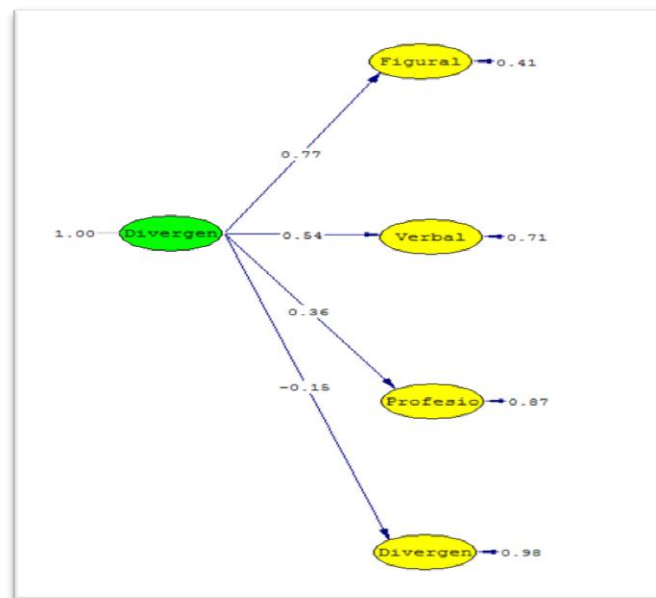


Figure 1. Construct of Instrument

Figure 2 shows the relationship between the latent variables with an observed variable measured is the loading factor of each component. There is a loading factor is negative. This means that the components of the divergent tests have probably not included in the component parts measure the ability to think creatively kindergarten teacher. Here are the results of loading factor in the assessment of the structural model of creative thinking kindergarten teacher described in Table 2 .

Table 2. The Decission for Construct Measurement

Part of instrument	Decission	Item
Convergent Figural	Accepted	2,4,8,10,11,12,13,16,17
	Rejected	1,3,5,6,7,9,14,15,18,19,20
Convergent Verbal	Accepted	1,2,3,4,5,7,8,11,12,14,18
	Rejected	6,9,10,13,15,16,17,19,20
Divergent	Accepted	1,2,3,4,5,6,7,8,9,10,11
	Rejected	12,13
Teacher Professionalism	Accepted	1,2,3,4,5,6,7,8
	Rejected	-

3. Instrument *Reability*

a. *Cronbach Alpha coefficients*

Calculation of the coefficient of reliability of the instrument basically using Cronbach Alpha formula . The level of reliability of the instrument is determined based on the coefficient owned. The higher the reliability coefficient , the higher the reliability of the instrument. The criteria used as the minimum limit of the coefficient of reliability in this evaluation was 0.65 . According Mehrens & Lehmann [14] , the level of reliability equal to or more than 0.65 , the instrument is quite good . Djemari Mardapi [15] , states that if the instrument has been assessed then repaired and subsequently assembled to be tested .

Table 3. Cronbach Alpha Coefficient

No.	Subtest	Reliabilitas (<i>Chronbach Alpha</i>)	Decision
1	Konvergen Figural	0,614	Cukup Baik
2	Konvergen Verbal	0,560	Kurang Baik
3	Divergen	0,898	Baik
4	Profesionalisme Guru TK	0,755	Baik

b. Composite Reliability

Composite reliability values or Average Variance Extracted (AVE) can be used to determine the reliability of each latent variable using the formula :

$$pc = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum_i var(\epsilon_i)} \dots\dots\dots [16]$$

Where λ_i is the component loading into indicators and $var(\epsilon_i) = 1 - \lambda_i^2$. The nature of the composite reliability is a closer approximation assuming the parameter estimates are accurate. Here are the results of calculation of composite reliability for each component.

Table 4. Reliability Coefficient

No.	Instruments	Composit Reliability	Decission
1	Convergent Figural	0,673	Good
2	Convergent Verbal	0,871	Good
3	Divergent	0,870	Good
4	Profesionalism	0,798	Good

Recapitulation result of composite reliability in Table 5 shows all the components have a good internal consistence that is above 0.65. It can be concluded that the composite reliability assessment instrument of kindergarten teacher creative thinking is good.

4. Study of Final Product

a. IRT Instrument Test assumptions of Kindergarten Teacher Professionalism

1) Unidimensional

Unidimensional test was conducted by factor analysis using SPSS 23. Before conducting factor analysis, it was conducted feasibility testing by using test analysis KMO - MSA and Barlett's test on each instrument. According to Anderson (1998, p.88) requirements for factor analysis is Kaiser - Meyer Olkin (KMO) - MSAU > 0.5 and significant barletnidimensi test means that every item tests only measure one 's ability. The results of the empirical analysis KMO - MSA value is equal to 0.826 or more than 0.5 and sig Barlett.s test is 0,000. Having in mind the total variance of 32.94 % in the first component can be interpreted this instrument to measure one aspect of the dominant eigen value is 3.624 means that the instrument developed to measure only one dimension of ability alone.

2) Analyzing Suitability Model

GPCM model test with the data is an analysis to see the characteristics of each question by looking at the parameter effect on each question. The criteria in the analysis are: 1) The value of Chi Square and probability in answering questions > 0.05, 2) Power difference (0-2), 3) The level of difficulty, and 4) Pseudoguessing (estimate).

3) The value of Chi-Square

Compatibility of this model can be seen by comparing the chi-squared (estimated at a table with a degree of autonomy and independence. The decision was taken when the item fits with the model GPCM is <table, whereas the item does not fit the model GPCM if> table. The estimation results on the value of Chi Square it can be concluded that whole grains are incompatible with the model GPCM ie 11 items. the items that match the model concluded GPCM means Ha rejected or empirical data in accordance with the model used. it can be concluded for teacher professionalism questionnaire instruments better suited wearing GPCM models.

4) Probability Item Values of Generalized Partial Credit Model

The first analysis is to look at the characteristics of items using GPCM. GPCM method is the development of Rasch models using two parameters, namely grains of different power and item difficulty were analyzed using Parscale software for Windows. Fit analysis assuming the probability of answering the questions is if the value of $P(\theta) < 0.05$ then the item is significant. Conversely, if the value of $P(\theta) > 0.05$ then the item is not significant. Recapitulation of the value of the probability of answering correctly. The analysis shows there are 2 items that do not fit with the model that is the number 1 and number 8. Based on the analysis performed GPCM models is analyzed by two parameters, namely the level of difficulty and different power.

5) Item Parameter Estimation

There are several criteria a good question based on the analysis IRT, namely:

- a) Power difference (a): 0 s / d 2
- b) The level of difficulty (b): -2 s / d 2

Based on analysis of 3 PL and criteria, there are several issues which are not in accordance with the following criteria:

- a) Power depending matter: all the questions that were analyzed have different power in the range of 0 s/d 2. depending Southwestern lowest for the question number 7 (0237) and different power is highest at about number 10 (1005).
- b) The level of difficulty of questions: about the difficulty level ranges from -2010 (No. 3) s / d -0.489 (No. 1). The whole matter has a difficulty level (-), it means that the instruments do not have the questionnaire in level of difficulty at all. Or it can be concluded all the items can be said to be easy to do.

b. Convergent Thinking IRT Instruments Test assumptions

1) Unidimensional

Unidimensional test conducted by factor analysis using SPSS 23 for windows. Before conducting factor analysis testing the feasibility of using test analysis KMO - MSA and Barlett's test on each instrument. According to Anderson et al . (1998, p.88) requirements for factor analysis is Kaiser - Meyer Olkin (KMO) - MSAU > 0.5 and significant barletnidimensi test means that every item tests only measure one 's ability. To test unidimesi by factor analysis. The results of the analysis KMO and Barlett, s less than 0.05. The results of the empirical analysis KMO - MSA value is equal to 0.672 or more than 0.5 and sig Barlett.s test is 0,000.

2) Analyzing Suitability Model

a) Chi Square Value and Item Probability

After a chi-square analysis it can be arranged recapitulation chi-square analysis suitability model as shown in Table 26 below. The results of the analysis in Table 26 below can be concluded that the model fits to 2PL and 3PL which has Chi Square smaller than χ^2 critical. And it can be seen that the model does not fit to the single parameter model (1PL) which has five items that do not fit with the model. After the probability analysis it can be arranged recapitulation model fit Chi-square analysis results of the analysis on the value of probability it can be concluded that the model matches the 3 PL model.

Kindergarten teacher convergent thinking instruments will provide information on the maximum test 0 at θ value of -0.125 and the test is suitable for capability scale approximately -3.5 up to -0.025. The error measurement in model 1 PL is 0.4 to 0.5. In 2PL model indicates maximum information about the function of -2.02 on θ 0.25 and the test is suitable for θ between -2.8 and 2.5. The error measurement is 2.1388. In the 3PL model, maximum test information is 0.4234 at θ is 1.5 and it is suitable for θ value between -0.8 to 3.9. The error measurement is 1.5368. Base on the three models above, the maximum information value verbal tests are on model 3 PL so that the error measurement of 3 PL model is smaller than in the other two models.

c. Divergent Thinking assumptions Test Instruments

1) Unidimensional

The empirical analysis results which KMO - MSA value is equal to 0.917 or more than 0.5 and sig Barlett.s test is 0,000. It can be concluded that all analysis results have been significant, which means the instrument are feasible to do factor analysis. The variance total in Table 35 is 59.1 % in the first component, it can be interpreted this instrument measures an aspect of the dominant eigen value of 4.72, it means that the instrument can be developed only to measure a capability dimension.

The Steep Sreeplot shows that there is a dominant component, which means that the teachers professionalism instrument measures only one factor or one dimension. In divergent capabilities instrument subtest there are several things to be measured, they are the original ideas ability (originality), elaboration and imagine imagination (abstractness).

2) Analyzing *Suitability Model*

a) Chi Square Value and Item Probability

The χ^2 values estimation describes, it shows that there are four items that χ^2 have a greater value than χ^2 critical, which means four items are match to GPCM models. While 4 point (2, 4, 5, and 6) has a $< \chi^2$ critical chi -square value, which means 4 grains are incompatible with the GPCM model.

b) Item Probability Value

Based on the analysis GPCM models which have been done, it was analyzed by two parameters, namely the difficulty level and different power. The Recapitulation shows that there are three items that do not fit the GPCM model are the number 1, 3 and 7.

3) Effectiveness *Developed Products*

The effectiveness of creative thinking Assessment are views from three aspects: 1) the accuracy or reliability of kindergarten teacher creative thinking abilities measurement results; 2) the ease of the product use; and 3) the use of product in the kindergarten program implementation. In the field, the empirical criteria developed by the reference or consideration of the promoter and expert (expert/specialist).

The criteria will be compared with the field data so that it shows the product effectiveness. In order to know the products effectiveness that have been developed, it was conducted by distributing questionnaires to the user in this case the head of the kindergarten.

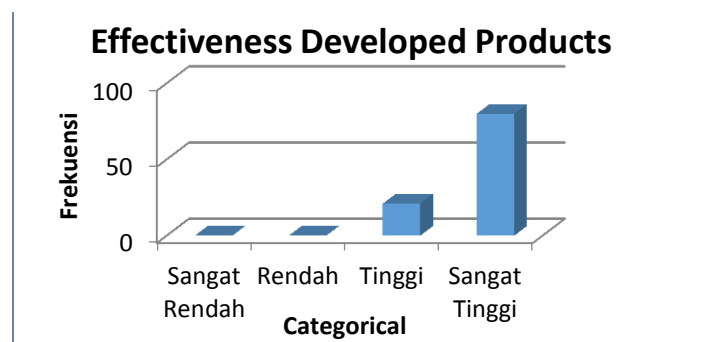


Figure 2. Effectiveness developed products

IV. CONCLUSION AND SUGGESTIONS

A. Conclusions

The kindergarten teachers' creative thinking constructive assessment instrument covered 3 components (teacher's professionalism, convergent thinking, and divergent thinking).

The characteristic development product is; a) on each instrument: (1) professionalism teacher instrument: (a) unidimensional, (b) suitable with the model GPCM (Generalized Partial Credit Model), (c) the optimal estimation TIF on theta -2.02; (2) divergent thinking instrument: (a) unidimensional, (b) suitable with the model GPCM, (c) the optimal estimation TIF on theta 0.25; (3) convergent thinking instrument: (a) unidimensional, (b) suitable with the model 2PL and 3PL, (c) the optimal estimation TIF on theta 1.5; b) the developed model was identical to the theory (H_0 is rejected), c) the developed product's effectiveness is categorised as high.

B. Suggestions

The Kindergarten teachers creative thinking skills assessment instrument can be used to measure the kindergarten teacher creative thinking ability by the Principal, Supervisor or City/Regency Education Department and selectors team of educator resources in kindergarten.

This product can be enhanced by developing test according to the relevant themes to early childhood education so that the instrument is more suitable for kindergarten teachers.

This product can be used as an early model of the kindergarten teachers creative thinking ability measurement by adding factors that are not included in the development of this research study.

REFERENCES

- [1] Bloomberg, M. (1973). Creativity. theory and research. Manufactured in the United States of America: United Printing Service.Inc.
- [2] Modul PLPG (Pendidikan dan Pelatihan Profesi Guru) Pendidikan Anak Usia Dini. (2013). Universitas Negeri Jakarta.
- [3] Petunjuk teknis penyelenggaraan taman kanak-kanak. (2011). Direktorat pembinaan PAUD. Dirjen PAUD, nonformal dan informal.
- [4] Angela, A. (2009). Early childhood education. society and culture. California: Sage Publitions.
- [5] Suara Merdeka. (2013). *Gaji guru honorer di bawah buruh*. Ditulis oleh K3-60. Koran Suara Merdeka 9 September 2013.
- [6] Jimmy, A. (2013). *Kemdigbud akui kualitas guru masih rendah*. Koran ANTARA. 27 September 2013.
- [7] Haryati, M. 2004. Model dan teknik penilaian pada tingkat satuan pendidikan. Jakarta: GP Press.
- [8] Glazer, E. (2009). Rephrasing the madness and creativity debate: What is the nature? Elsevier-Personality and Individual Differences. Ed 46.
- [9] Belkhasas. (2010). A Dessertation of Doctoral: Creative teaching to increase students' learning and achievement the case of english teachers - University of Constantine. Published.
- [10] Sternberg, R.J. (1999). Handbook of creativity. Australia: Cambridge University Press.
- [11] Torrance, E.P. Future careers for gifted and talented students gifted child, Quarterly 20: 142-156. 1976.
- [12] Branch Robert M. (2009). Instructional Design: The ADDIE Approach. Athens. USA: Springer Publisher.
- [13] Ghozali, I. (2014). Structural equation modelling: Teori, konsep, aplikasi dengan Program Lisrel 9.10. Semarang: Badan Penerbit Universitas Diponegoro.
- [14] Mehrens, W.A. & Lehmann, L.J. (1973). Meassurement and evaluation in educational and psychology. New York: Hold, Rinehart and Winstons Company.
- [15] Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Jogjakarta.
- Anderson, R.E. (1998). Multivariate data analysis. USA. Prentice-Hall. International.Inc.

Indonesian Adaptation Scale for Job Content Questionnaire (JCQ)

Sandra Jati Purwantari

Faculty of Psychology (UIN Syarif Hidayatullah Jakarta)
sandra20190@gmail.com

Abstract - JCQ was used to measure work stress and based on the model of job demand-control by Karasek, has been widely used and translated into the 23 official languages, but not yet officially available in Indonesian. This research was aimed to analyse the construct validity of the Indonesian Version of the JCQ among bank employees in Jakarta (n=104; aged 19-45 years) using confirmatory factor analysis (CFA). Adaptation method was forward translation (English-Indonesian) and back translation (Indonesian-English). This procedure was conducted to ensure the face validity of the questionnaire based on International Test Commission guidelines. The results indicate that the confirmatory factor analysis showed items could explain the three theoretical dimensions or construct of Karasek's demand-control-social support model. In conclusion, the results of the validation study suggested that the adaptation scale of the JCQ is valid for assessing the work stress. Further analysis are necessary to evaluate the stability and concurrent validity of the JCQ.

Keywords: JCQ, Karasek, adaptation scale, confirmatory factor analysis, work stress

I. INTRODUCTION

Nowadays, workplaces are changing and competitive increase the level of stress among workers and managers. For example, a survey of US workers found that 46% felt their job as stressful and 34% percent think seriously to quit their job 12 months earlier because of stress in the workplace (Schellhardt, 1986). One cause stress is due to the work load (job content) are large and have to work longer because of downsizing at their company. We read a survey where employers complain about the stress that is created in an effort to compensate for the work and family responsibilities.

The American Psychological Association's annual stress survey, in 2012, claimed that 65% of the work as the main source of stress, only about 37% of Americans surveyed said they did an excellent job of managing stress or excellent. The demands of the job is psychological stressors in the work environment. This includes factors such as: the work load (job content), the level of distraction, time pressure, conflicting demands, the reaction time is needed, the pace of work, the proportion of work performed under pressure, the number of jobs, the level of concentration required, and slows jobs caused by the need to wait for others.

Work stress is a condition that affects the emotional tension, thought processes, and physical condition, if the stress is too large it can threaten a person's ability in dealing with the environment (Davis and Newstrom, 1985). Another understanding of Holroyd and Lazarus (1982) which requires assessment of psychological stress that the demands of the environment and / or internal exceeding individual resources to manage it. The point of stress that because there is an imbalance between the ability of individuals with existing demands.

The demand-control model was first formulated by Karasek in 1979 (Karasek, 1979), it has been shown that 3 dimension - job control, psychological demand, and social support - are important determinants of work stress, which in turn have significant effects on health, in particular cardiovascular health (Theorell and Karasek, 1996). The validity of the JCQ in various languages has been assessed in some recent works using the English, Dutch, and Japanese versions (Karasek et al, 1998).

Have been many studies that measure reliability and validity job content questionnaire, by measuring the factor validity of using explanatory factor analysis (EFA) or confirmatory factor

analysis (CFA). Storms (2001) showed that the JCQ adaptation into German is a solid measurement instrument based on the results of measurement reliability and validity. Adaptation JCQ into the Thai language version, which uses EFA and CFA analysis also showed reliability and validity that can be relied upon, even if research conducted on two versions of JCQ are 22 items and 45 items (Phakthongsuk & Apakupakul, 2008).

Literature on Adaptation JCQ that have been found in research that is to the version in Persian (Choobineh, Ghaem, and Ahmedinejad, 2009), the language version Mexico (George, Jimenez, Munoz, and Uribe, 2011), the language version Malaysia (Edimansyah, Rusli, Naing & Mazalisah, 2006), Korean language version (Eum et.al., 2007), the Polish language version (Zreda & Bedynska, 2014), that all this research uses only EFA analysis, and all of them showed good results in the reliability and validity results adaptation. Based on the literature, the most recent study of adaptation JCQ is adaptation in the Greek version, using 29 items, uses a forward-back translation method translation, and analysis using the EFA and CFA. The results showed that the Greek version of the JCQ has the reliability and validity to Investigate psychosocial job characteristics among health workers Greece (Alexopoulos, Argyriou, Bournas & Bakoyannis, 2015).

JCQ adaptation research into language Belgium conducted on a very large number of samples reached 21419 workers in Belgium, which is also used confirmatory factor analysis, found that there was indeed a correlation between factors (Pelfrene et.al., 2001). De Araujo & Karasek (2008), early researchers who formulate JCQ, in doing research into adaptation JCQ language version of Brazil, came to the conclusion that JCQ findings can be used in studies Carried out in developing countries and in informal situations in the which jobs are common. because Indonesia is also developing the adaptation research into Indonesian version can be done based on Karasek research previously conducted in Brazil. There is an urgent need to initiate studies of work stress using a valid and reliable JCQ as an instrument designed to assess the psychological work environment in Indonesia. The purpose of this study was to assess construct validity of the Indonesian version of the JCQ in Jakarta, using confirmatory factor analysis.

A. Job Content Questionnaire

In the measurement of work-related stress, using measurement scale Job Content Questionnaire (Karasek, 1985) as a demand-control model application. Demand-control model of Karasek on job stress has a major influence on the design of work and occupational health because it is practical and can be tested (Jones & Bright, 2001). JCQ questionnaire has been already widely used in various countries, which now has been translated into 23 official languages in the JCQ Center, but has not yet officially available in Indonesian.

In the model-control demands, workplace stress is a function of how a person's work was demanding and how much control (wisdom, authority or decision etc.) owned more than their own responsibilities. The demands of the job is defined as a psychological stressor, such as the need to work hard and fast, the amount of work and a lot of time is limited. According to Karasek (1979) in the job demands-job control model, the result of stress such as problems on mental and physical health, occurs when the work simultaneously there is a very high job demands but control of the work is very low.

Karasek provide four categories of job stress, as follows: High-Strain Job (stress will happen if simultaneously if someone faced with high job demands but have no control against lowly jobs), Low-Strain Job (in conditions where by increasing control over the demands of the job but the job remains low, it will create a helplessness, which is referred to as "low-strain job"), Active Job (the condition of the occurrence of a positive outcome, such as motivation, learning, and improvement in health will arise when a person is in a condition where has the job demands and job control over the equally high), and Passive Job (condition where a person has a job demands and low control over their work. This will increase the level of job dissatisfaction, and over time they will not be able to make a decision, not being able to face challenges and solve problems).

B. Dimensions of Job Content Questionnaire

1. Job Control

Job Control or control of the work, for the first time included as a variable that causes stress, a theory created by Karasek (1979). Control over work (control) has two main components (Karasek & Theorell, 1990), namely: power / authority of the workers to make decisions in his

work (decision authority) and flexibility for workers to use his expertise in the work (skill discretion).

2. Psychological Demands

In the model is defined as psychological demands of the job demands, such as the need to work hard and fast, the amount of work and a lot of time is limited. The demands of the job in this case is analogous to the workload. However, that needs to be emphasized is that the demands of work here is a requirement of the psychological rather than the demands physically, although the need for speed and complexity of work requires physical abilities that can have an impact on fatigue, but the stress outcome predicted by this model with regard to the psychological impact on work-burden associated with the need to set the pace in the work and the consequences of failure to complete the work.

3. Social Support

Social support is feeling comfortable, cared for, respected, or receive help from other people or groups. Someone will receive social support depends on the amount, composition, proximity and frequency of contact person with the social network. Social attachment and relationships with others that lasts long been accepted as an aspect of emotional satisfaction in life. According to Karasek (1998), there are two dimensions that determines the strength of support in the work include "supervisor is the support received by employees from their superiors" and "work colleagues are the support received by employees from their colleagues".

II. METHOD

A. Study Design and Sample Size

Dissemination of data was done by using self-assessment report. Researchers distributed questionnaires that have been adapted to one branch of BNI Syariah based in Jakarta. Criteria for the study population defined by the researchers is to employees both men and women aged 19-45 years and fill out a questionnaire to complete. The sampling technique used in this research is purposive sampling technique for the respondent to be elected to the sample on the basis of the researchers themselves. Respondents in this research were 130 people, but 26 of whom did not fill out a questionnaire to complete and therefore do not meet the criteria. So that the sample in this study amounted to 104 people (women 61,5% and man 38.5%).

B. Questionnaire

Twenty items taken from JCQ adaptation research into language version Malaysia (Edimansyah, Rusli, Naing & Mazalisah, 2006), which refers to the full recommendation JCQ consists of 49 items. These items constituted a minimum set of questions for three major scales of the JCQ; Job Control (9 items), Psychological Demands (5 items), and Social Support (6 items). Items in the scales were recorded using the Likert scale, ranging from 1 (strongly disagree) to 4 (strongly agree).

C. Linguistic Validation

The first step was translation to ensure the linguistic validity of the questionnaire in the new language. The 20-item JCQ was translated into Indonesian by a qualified translators. The method chosen is a forward-back translation. The first translator is a person skilled English / language origin of the scale that will be adapted. The second translator also have to master The English and proficient content such as undergraduate psychology or a psychology professor. After getting the results of the translators one (T1) and two translators (T2), the researchers then studied the results of T1 and T2. If found differences between both the translation results, researchers will choose which one has the most meaning in accordance with the original questions. In this process the researchers also consider cultural factors in choosing the translation. Then performed the translation back to the initial language scale, by a qualified English translator (T3), blind to the original English questions version (backward translation).

D. Statistical Analysis

The construct validity of the instrument was examined using Confirmatory Factor Analysis, using Lisrel 8.7. The first step in the analysis process was to test the measurement model. The testing of the measurement models involved specifying the degree to which the observed variables (indicators) defined a construct or latent variable through the use of CFA. The relationship between observed and the latent variable (Lambdas) are viewed as "factor loadings"

and interpreted as validity coefficient. A significant Lambda coefficient has a t-value equal to or greater than 1.96.

The next step was to test the structural model. Structural equation models are often diagrammed by a path model in which the hypothesized factors are viewed as latent variables. Lisrel provides a number of fit indicators that test the fit between the data and the theoretical model via the chi-square test method (p-value > .01 is evidence that the theoretical model fits the data). The root mean squared error of approximation (RMSEA) focused on the discrepancy between the sample covariance matrix and the hypothesized model covariance matrix but did also account for the degree of freedom.

III. RESULT

The following results of the analysis of 3-dimensional consisting of 20 items of questions using CFA. Nine items for job control existing items are unidimensional, means it only measures the dimensions of job control. From the analysis conducted by the CFA model of a factor obtained model fit, with Chi-square = 32.78, df = 22, P-value = 0.06506, and the value of RMSEA = 0.000. This means that all significant items in this dimension that measures only one factor-dimensional job control. however, based on Table 1, only the t-value for the item number 6 which is not eligible significant (significant = t-value > 1.96). So, item number 6 should be dropped or re-tested for further research with larger sample sizes.

Table 1. Dimensional Factor Loading Items - Job Control

Item	Lambda Coefficient	Standard Error	t-value	Significant
1	0.72	0.09	7.90	V
2*	0.26	0.10	2.51	V
3	0.79	0.09	8.96	V
4	0.32	0.10	3.14	V
5	0.76	0.09	8.76	V
6*	0.02	0.11	0.15	X
7	0.48	0.10	4.86	V
8	0.53	0.10	5.46	V
9	0.70	0.09	7.52	V

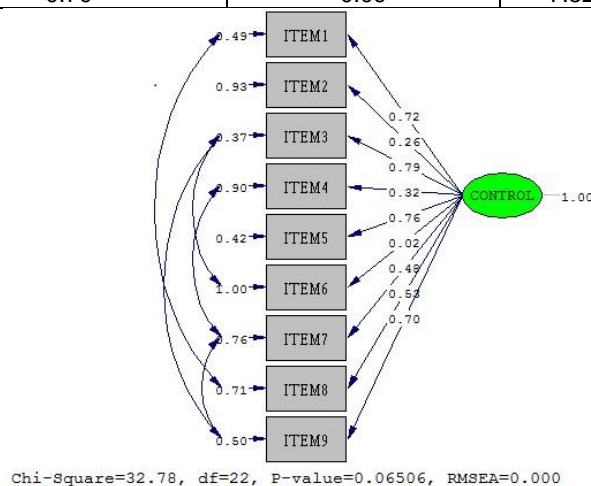
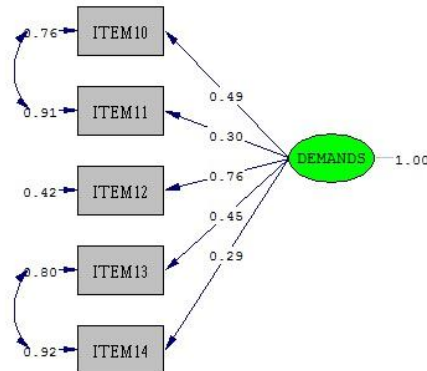


Figure 1. Job Control Path Model

Five items for psychological demands dimension are unidimensional, means it only measures the dimensions of psychological demands. From the analysis conducted by the CFA model of a factor obtained model fit, with Chi-square = 4.55, df = 3, P-value = 0.20775 and the value of RMSEA = 0.000. This means that all significant items in this dimension, that is only one factor to measure the dimensions of psychological demands. Based on table 2, t-value for all items in the dimension of the psychological demands significant qualify so that no item is dropped.

Table 2. Dimensional Factor Loading Items - Psychological Demands

Item	Lambda Coefficient	Standard Error	t-value	Significant
10	0.49	0.13	3.66	V
11	0.30	0.13	2.31	V
12	0.76	0.17	4.43	V
13	0.45	0.13	3.46	V
14	0.29	0.13	2.31	V



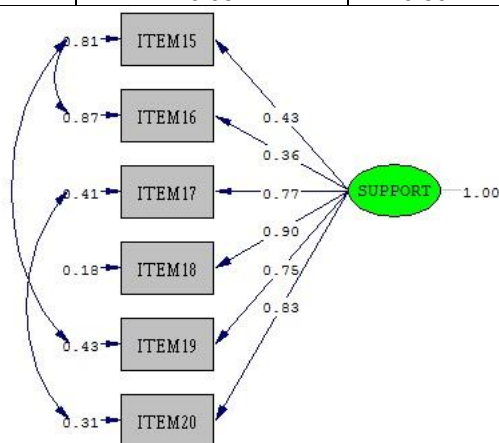
Chi-Square=4.55, df=3, P-value=0.20775, RMSEA=0.000

Table 3. Psychological Demand Path Model

Six existing items are unidimensional for social support dimension, means it only measures the dimensions of social support. From the analysis conducted by the CFA model of a factor obtained model fit, with Chi-square = 7.86, df = 6, P-value = 0.24854 and the value of RMSEA = 0.000. This means that all significant items in this dimension, that is only one factor to measure the dimensions of social support. Based on table 2, t-value for all items in the dimension of the job support significant qualify so that no item is dropped.

Table 4. Dimensional Factor Loading Items - Social Support

Item	Lambda Coefficient	Standard Error	t-value	Significant
15	0.43	0.10	4.40	V
16	0.36	0.10	3.63	V
17	0.77	0.09	8.84	V
18	0.90	0.08	11.52	V
19	0.75	0.09	8.80	V
20	0.83	0.08	9.89	V



Chi-Square=7.86, df=6, P-value=0.24854, RMSEA=0.000

Figure 2. Social Support Path Model

The result was p-value for all the dimensions can be accepted (p-value>0.05) or classified as model fit. Lambda coefficient all items are positive and the t-value for the coefficient of factor

loadings of all items (except one item in Job Control dimension) are significant (t -value > 1.96) and indeed measure what it should be.

IV. DISCUSSION

The present study is an attempt to validate the psychometric properties especially measuring the factor validity using Confirmatory Factor Analysis (CFA) of the Indonesian version of selected scales of the Job Content Questionnaire (JCQ). The results of confirmatory factor analysis showed that items in questionnaire were clearly associated with the dimension of job control, psychological demands, and social support, also can be classified as a model fit.

In the linguistic validation stage of my study, using forward-back translation method, the main objective was to obtain a conceptual equivalence between the original and translated version of the JCQ. The translation of the JCQ into Indonesian was carried out within the framework of a standardized method of adaptation based on International Test Commission guidelines, that there was no major ambiguity in the translated version and, with few exceptions, the participants understood the questions.

A major limitation of this study was that the sample size was rather small, so that there is one item in job control dimension that do not qualify for significant likely due to the number of samples that need further research. Other limitation of this study was that the subjects were from a certain occupational group (i.e. Bank employees) and the study population or sample did not include a variety of occupations. The Indonesian version of the JCQ should be applied to workers in other occupations in future studies.

In conclusion, this study demonstrated that items in Indonesian version of JCQ were valid for assessing the psychosocial work condition of Indonesian workers, although further studies will be needed to improve the Job Control (especially item number 6 that mentioned above). It should be emphasized that the current study offers only preliminary findings in Indonesia. Complete validation demands the application of the psychometric scales in larger samples and more sophisticated methodology.

ACKNOWLEDGMENT

The author wish to thank Dr. Yunita Faella Nisa, M.Psi for her assistance and suggestion in development of the Indonesian version of the JCQ. The author also wish to thank Mr. Teguh and Mrs. Jane for their assistance in forward translation and Mrs. Melly for her assistance in backward translation. Finally, the author would like to express sincere gratitude to UIN Syarif Hidayatullah Jakarta for their support.

REFERENCES

- [1] Theorell T., & Karasek R. (1996). Current issues relating to psychosocial 2. job strain and cardiovascular disease research. *Journal of Occupational Health Psychology*, 1:9–26.
- [2] Karasek R. (1998). The Job Content Questionnaire (JCQ): an 9. instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology*, 3:322–355.
- [3] Karasek R. (1985). *Job Content Questionnaire and User's Guide*. Lowell, Massachusetts: Department of Work Environment, University of Massachusetts
- [4] Evangelos C., Alexopoulos, Argyriou E., Bournas V., & Bakoyannis. (2015). Reliability and Validity of the Greek Version of the Job Content Questionnaire. In *Greek Health Care Workers. Safety and Health at Work Article*, volume 6, Issue 3, pages 233–239
- [5] Storms G, Casaer S, de Wit R, van den Bberg O, & Moens G. (2001) A psychometric evaluation of a Dutch version of the job content questionnaire and of a short direct questioning procedure. *Work Stress Journal*, 15:131–43.
- [6] Pelfrene E, Vlerick P, Mak R, de Smet P, Kornitzer M, & de Backer G. (2001) Scale reliability and validity of the Karasek «Job-Demand-Control-Support model in the Belstress study. *Work Stress* 2001, 15:297–313.
- [7] Karasek, R. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, pages 285–308.
- [8] Hymowitz, C. & Schellhardt, T.D. (1986) The glass-ceiling: Why women can't seem to break the invisible barrier that blocks them from top jobs. *The Wall Street Journal*
- [9] Davis, K., & Newstrom. (1985). *Human Behavior at Work : Organization Behavior*, McGraw Hill Book Company New York.
- [10] Holroyd, K. A., & Lazarus, R. S. (1982). Stress, coping, and somatic adaptation. In L. Goldberger & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects*. New York: Free Press.
- [11] Jones F., & Bright J. (2001). *Stress: Myth, theory and research*. Harlow, England: Prentice-Hall.
- [12] Eum KD, Li J, Jhun HJ, Park JT, Tak SW, & Karasek R. (2007) Psychometric properties of the Korean version of the job content questionnaire: data from health care workers. *Int Arch Occup Environ Health*, 80:497–504.

- [13] Edimansyah B.A., Rusli B.N., Naing L., & Mazalisah M. (2006). Realibility and Construct Validity of the Malay Version of the Job Content Questionnaire. Southeast Asian Journal of Tropical Medicine and Public Health
- [14] George F.V.I., Jimenez B.M., Munoz A.R., Uribe J.V. (2011). Factorial Validity of the Job Expectations Questionnaire in a Sample of Mexican Workers. The Spanish Journal of Psychology ,Vol. 14, No. 2.
- [15] De Araujo T.M., & Karasek R. (2008). Validity and Reliability Of The Job Content Questionnaire In Formal and Informal Jobs In Brazil. Scandinavian Journal of Work, Environment & Health, pg. 52
- [16] Choobineh A.,Ghaem H., & Ahmadinejad P.(2009). Validity and reliability of the Persian (Farsi) version of the Job Content Questionnaire: a study among hospital nurses. Eastern Mediterranean Health Journal, Vol.17, no.4
- [17] Phakthongsuk P & Apakupakul N.(2008). Psychometric Properties Of The Thai Version Of The 22-Item and 45-Item Karasek Job Content Questionnaire. International Journal of Occupational Medicine and Environmental Health, 21(4):331 – 344
- [18] American Psychological Association (2012, January 26). Stress in America™: Our Health at Risk, pages 31-35

DEVELOPMENT OF ASSESSMENT INSTRUMENTS OF ART PAINTING PRODUCTION INTEGRATED WITH CHARACTER FOR ASSESSING LEARNERS' FIELD WORK PRACTICE IN VOCATIONAL HIGH SCHOOL

Trie Hartiti Retnowati, Djemari Mardapi, Bambang Prihadi
Yogyakarta State University, trihartiti54@gmail.com, 082121161604

Abstract- The research objective is to develop a set of instruments and scoring guides of nontest assessment for assessing art painting production that are integrated with character assessment in an effort to prepare students of Vocational High School in the face of the world of work. Based on early studies, there are not currently available such instruments and scoring guides as a tool to prepare vocational high school students towards the world of work. This study used the research design and development of Borg and Gall including the following procedures: (1) initial study to find research findings related to the process that will be developed, (2) development of a process based on the findings of the research, (3) field testing in a real situation in which the process will be used, and (4) improvement of instrument products based on the weaknesses found in the field testing. This research resulted in a set of non-test instruments of art painting production integrated with character assessment and scoring guides to assess the creative process and product of vocational high school learners in creating works of painting which can be used to determine their job readiness. In accordance with the existing regulations, non-test assessment is a part of the assessment of learning outcomes of the field work practice in vocational high school.

Key words: *assessment, non-test instrument, art production, character*

I. INTRODUCTION

A. Background

Education in Indonesia at this time seems more concerned with the mastery of knowledge dimension and tends to ignore the education of values. Consequently the Indonesian people need to generate the commitment and do good character education in families, schools, and communities. This condition and the current situation demands character education that needs to be transformed from an early age, namely from early childhood education, elementary education, secondary education, and higher education in a holistic and sustainable way.

Character education is an investment in cultural values that builds character, morality, and public personality which is taken in a long, continuous, intense, constant, and consistent way. Thus, character education give the students science, knowledge, and cultural practices which is oriented on the values of an ideal life, which are based on local culture (local wisdom) and foreign cultures (Tranggono, 2010: 27)

Our national commitment on the need for character education imperatively stated in Article 3 of Law No. 20 of 2003 on National Education System. It is stated in the Act that "The national education serves to develop the abilities and build character and dignified civilization of the nation in the context of the educating the life of the nation, is aimed at developing students' potentials in order to become a man of faith and fear of God Almighty, noble, healthy, knowledgeable, capable, creative, independent, and become democratic and responsible citizens." If observed, most of the learners' potentials to develop are closely related to character. Thus, the function and purpose of national education, as mandated by the Act, is not only to develop their knowledge and skills but also the character of the students.

The implementation of character education program is integrated into all school subjects, including arts. This principle brings students to learn through the process of thinking, being, and doing. Art education is closely related to character education because of their uniqueness, meaningfulness, and usefulness for developmental needs. Through art lessons the students are given aesthetic experience in the form of art production and art appreciation which includes moral values. Thus, the character building of children with art education in schools can be implemented through such activities in the classroom, so that the role of art education in schools can be as a means to build the learners' character in order to prepare the 2045 generation.

Vocational High Schools, in accordance with their characteristics, prepare their students to enter the world of work, so that the relevance of their education holds the key importance. Relevance must be interpreted as conformity of competences learned by students with the demands of jobs that will be entered after graduation. Thus, vocational education should be able to anticipate the demands far ahead, starting from the current educational program designed. The curriculum of the vocational education leads to multi skills with strong basic skills, emphasizes the intellectual skills, adaptability and self-training, as well as friendly attitudes towards the local environment (Samani, 1995; End, 1997). Basic skills encompass life skills such as faith and morality, cooperative skills, the ability to explore and process information in facing the situation at hand, problem-solving skills, and the ability to communicate. These abilities are in the affective domain, including characters needed by learners for doing their jobs and living in a society. Thus, it is important to develop character in the learning process. The affective domain is related to attitude, then the measurement to use are non-test instruments. So far, there have been no such instruments and guidelines for assessing the learners' character in vocational high schools

In 2011 The Center of Book and Curriculum and the Ministry of Education and Culture issued general guidelines for the implementation of Character Education for schools. In the same year The Junior High School Directorate Development, Directorate of Basic Education and Culture, published the character education operational guidelines to help junior high school teachers implement the character education which are integrated in the learning process for all lessons. Both books are aimed to help teachers implement character education, but they have no instruction how to carry out the assessment of the students' character development operationally. Although the syllabi developed already contain character items to integrate in the learning process, they do not include operational guidelines for teachers to carry out the assessment of the character development of students. Given the importance of character assessment integrated in learning art, especially painting creation, and the lack of instruments and operational guidelines properly developed, this research is aimed at developing an assessment instrument of painting production integrated with the character, especially in the job training of vocational high school students.

B. Formulation of Research Problems

The research problems are formulated as follows:

1. How to develop instruments for accessing painting production integrated with character assessment for the job training of vocational high schools students?
2. How to develop the guidelines to use the assessment instruments for accessing painting production integrated with character assessment for the job training of vocational high schools students?

C. Research Purposes

The purposes of this study are as follows:

1. To develop instruments for accessing painting production integrated with character assessment for the job training of vocational high schools students?
2. To develop the guidelines to use the assessment instruments for accessing painting production integrated with character assessment for the job training of vocational high schools students?

D. Research Significance

The benefits of this study are as follows:

1. To provide the institution where the job training is undertaken with non-test instruments and guidelines in assessing students' performance relating to character.

2. To assist vocational high school teachers in assessing the students' readiness for job of painting production.
3. Provide vocational high school teachers with proper guidelines for assessing the students' performance in painting production for job.

II. METHODS

This study is a research and development which comprises such steps as follows: (1) the initial study to find research findings related to the product to be developed, (2) the development of products based on the research findings and a literature review, (3) field testing in a real situation in which the product will be used, and (4) improvement of the product based on the weaknesses found in the field test (Borg and Gall, 1983).

The research was conducted with the stages as presented in Figure 1. Firstly, the literature review and the results of previous studies to support the product to be developed i.e. a set of non-test instruments and its implementation guidelines. Secondly, the development of the instruments which comprises the development of the instrument prototype including the constructs, descriptors, indicators, blueprints, and usage guidelines. Thirdly, the Focus Group Discussion (FGD) with such stakeholders as the vocational high school teachers, studio owners or institutions where the job training is conducted, and art expert.

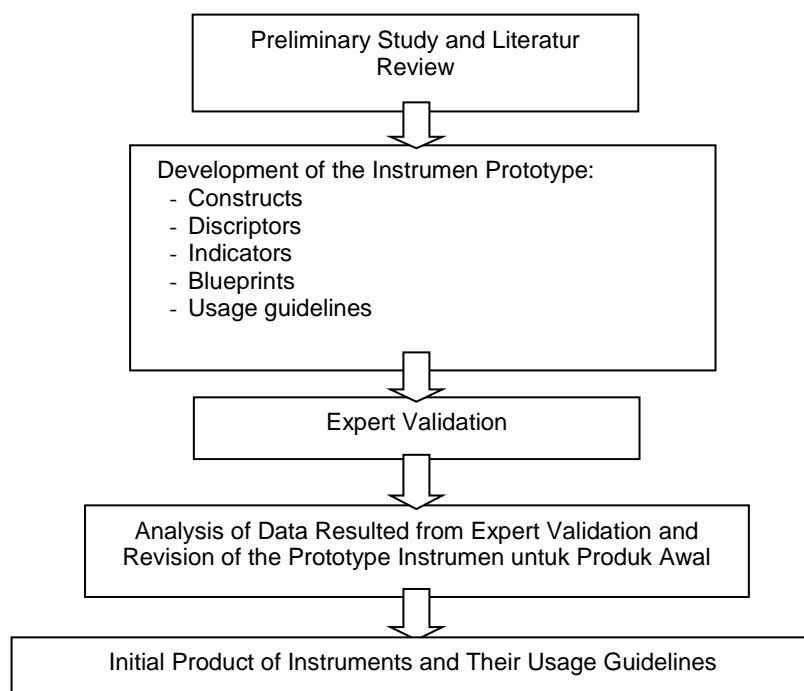


Figure 1. Development of Non-Test Instruments for Painting Production Assessment Integrated with Character Assessment

Thirdly, based on the results of the FGD, the assessment blueprints were discussed and revised and then validated by art education experts from Department of Art Education Faculty of Languages and Arts, Yogyakarta State University. Fourthly, analysis of the validation results and subsequent revision of the instrument prototypes and their usage guidelines as the products of this research.

III. RESULTS AND DISCUSSION

The initial step of the development of assessment instruments of painting production integrated with character assessment for the job training in vocational school was reviewing the relevant theories. The next step was identifying non-test instruments for assessing the painting production stages namely the preparation, the process of painting, and the product of painting.

The assessment techniques used for these steps were observation, peer evaluation, and self-assessment respectively which were equipped with rubrics.

In the initial stage, the development of instruments was done by examining the theories underlying the constructs to be assessed so that the instruments would measure what were supposed to be measured. The constructs were then translated into blueprints which included dimensions, descriptions, and indicators. The blueprints were further discussed in Focus Group Discussion (FGD) involving the fine arts teachers of vocational high school in Yogyakarta, art studio owners, and art experts. The blueprints were then revised based on the results of the FGD and validated by art education experts from Department of Art Education, Yogyakarta State University.

The blueprint of the preparation step comprised such dimensions as follows: mental and physical readiness of the student, readiness and completeness of painting tools and safety equipment, responsiveness and proactive attitude as well as student's initial knowledge initial skills in painting. This blueprint is presented in Table 1.

Table 1. Blueprint of Preparation Step

Number.	Dimension	Descriptors	Indicator
1	Student's mental readiness	The student's positive perception and attitude toward the job training	The student concentrate on and is confident and eager to undertake the job training
2	Student's physical readiness	The student's physical health to undertake the job training	The student has good physical conditions to undertake the job training.
3	Readiness and completeness of painting tools	Availability of painting materials and tools by the student	The student provides canvases, oil painting, brushes, pallet, easel, painting oil, turpentine, pastel, water color, poster color, acrylic, cloth, and painting frame.
4	Readiness and completeness of safety equipment	Availability of safety tools used to paint by the student	The student provides cloth, masker, and working clothes.
5	Responsiveness	The student's responsiveness to follow instructions	The student quickly follows the instructions given by the teacher and the studio owner (painter).
6	Proactiveness	The student's proactiveness to follow instructions	The student does things before the teacher and the studio owner (painter) give instructions.
7	Initial knowledge	The student's acquisition of method and procedures of painting	The student knows about method and procedures of how to paint pictures creatively.
8	Initial skills	The student's acquisition of skills in painting	The student is able to paint pictures using materials and tools creatively.

The technique used to assess the students' preparation for painting was observation. The instrument used was in the form of observation sheet as presented in Figure 3. This instrument is equipped with a rubric (Table 2) as a guide for the instructor to do the assessment.

OBSERVATION SHEET PREPARATION FOR PAINTING						
Name of student		:				
Name of School:		:				
Teacher (adviser)		:				
		Name of Studio				
Owner (Painter)		:				
Studio Address		:				
<p>Based on the observation that you do, give an assessment of the performance of students in the preparation by giving a check mark (✓) in the range of scores in accordance with the assessment rubric.</p>						
No.	Dimension	Range of score				Score
		1	2	3	4	
1	Student's mental readiness					
2	Student's physical readiness					
3	Readiness and completeness of painting tools					
4	Readiness and completeness of safety equipment					
5	Responsiveness					
6	Proactiveness					
7	Initial knowledge					
8	Initial Skills					
Sum of scores						

Instructor/Studio Owner

Figure 2. Observation Sheet for Assessing Preparation Step

Table 2. Rubric for Assessing Preparation Step

No.	Dimension	Range of Score			
		1	2	3	4
1	Student's mental readiness	The student shows no concentration on and shows no confidence and eagerness to do the job training.	The student shows lack of concentration on and show less confidence and eagerness to do the job training.	The student fairly shows concentration on and show confidence and eagerness to do the job training.	The student shows much concentration on and show much confidence and eagerness to do the job training.
2	Student's physical readiness	The student has poor health conditions to do the job training.	The student has rather poor health conditions to do the job training.	The student fairly has health conditions to do the job training.	The student has good health conditions to do the job training.
3	Readiness and completeness of painting tools	The student does not provides materials and tools for painting (canvases, oil painting, brushes, pallet, easel, panting oil, turpentine, pastel, water color, poster color, acrylic, cloth, and painting frame).	The student provide less materials and tools for painting (canvases, oil painting, brushes, pallet, easel, panting oil, turpentine, pastel, water color, poster color, acrylic, cloth, and painting frame).	The student provides enough materials and tools for painting (canvases, oil painting, brushes, pallet, easel, panting oil, turpentine, pastel, water color, poster color, acrylic, cloth, and painting frame).	The student provides all materials and tools for painting (canvases, oil painting, brushes, pallet, easel, panting oil, turpentine, pastel, water color, poster color, acrylic, cloth, and painting frame).
4	Readiness and completeness of safety equipment	The student never provides cloth, masker, and	The student sometimes provides cloth,	The student often provide cloth, masker, and	The student always provides cloth, masker, and

No.	Dimension	Range of Score			
		1	2	3	4
		working clothes.	masker, and working clothes.	working clothes.	working clothes.
5	Responsiveness	The student never follows the instructions given by the teacher and the studio owner (painter) quickly.	The student sometimes follows the instructions given by the teacher and the studio owner (painter) quickly.	The student often follows the instructions given by the teacher and the studio owner (painter) quickly.	The student always follows the instructions given by the teacher and the studio owner (painter) quickly.
6	Proactiveness	The student never does things before the teacher and the studio owner (painter) give instructions.	The student sometimes does things before the teacher and the studio owner (painter) give instructions.	The student often does things before the teacher and the studio owner (painter) give instructions.	The student always does things before the teacher and the studio owner (painter) give instructions.
7	Initial knowledge	The student knows less about methods and procedures of how to paint pictures creatively.	The student knows little about methods and procedures of how to paint pictures creatively.	The student knows some about methods and procedures of how to paint pictures creatively.	The student knows much about methods and procedures of how to paint pictures creatively.
8	Initial Skills	The student has less ability to paint pictures using materials and tools creatively.	The student has little ability to paint pictures using materials and tools creatively.	The student has some ability to paint pictures using materials and tools creatively.	The student has much ability to paint pictures using materials and tools creatively.

The result of the observation of the student's preparation was then graded using the following formula:

$$\text{Grade} = \frac{\text{Sum of scores acquired}}{\text{Maximum score}} \times 100$$

The second Instrument was used to assess the process of painting. The students' performance and attitudes during the painting process comprises such dimensions as follows: (1) the development of ideas, (2) the expression of the idea through the medium, (3) the usage of time, and (4) disiplin. This instrument was made in the form of inter-rater assesment. The dimensions, descriptors, and indicators for this instrument is presented in Table 3. The instrument to assess the process of painting is presented in Figure 4.

Table 3. Instrumen Penilaian Proses Prakerin Seni Lukis

No.	Dimensi	Descriptors	Indicators
1.	Development of idea	Speed and creativity in developing ideas for painting.	The student develop his ideas to paint pictures quickly and creatively.
2	Expression of the idea through the medium	The effectiveness of using painting materials and tools.	The student uses painting materials and tools in accordance with the right procedures and his expression.
3	Pemanfatan waktu	The efficiency of using time.	The student does his painting smoothly and never wastes time.
4	Discipline	Undertaking tasks in job training accordance with the rules and regulations	The student undertakes tasks in job training accordance with the rules and regulations by the school and the studio owner (painter).

INSTRUMENT OF INTER-RATER ASSESSMENT
PROCESS OF UNDERTAKING TASKS IN THE JOB TRAINING

Name of student :
 Name of School :
 Teacher (adviser) :
 Name of Studio :
 Owner (Painter) :
 Studio Address :

Give an assessment of the performance of your fellow student in undertaking the job training by giving a check mark (√) in the column "Yes" or "No" according to your own opinion.

Pernyataan	Yes	No	Score
According to my observation, the student above:			
1. develops his ideas quickly.			
2. develops his ideas creatively.			
3. develops his composition quickly.			
4. develops his composition creatively.			
5. uses materials and tools in accordance with procedures smoothly.			
6. uses materials and tools in accordance with expression smoothly.			
7. does his painting diligently.			
8. uses materials efficiently.			
9. uses effective steps in doing painting.			
10. undertakes tasks of the job training in accordance with the adviser.			
11. undertakes tasks of the job training in accordance with the schedule.			
Sum of scores			

Rater:

Name of School :
 Registration number :
 Signature :

Figure 3. Inter-rater Assessment Process of Undertaking tasks in the Job Training

The result of the observation of the student's preparation was then graded using the formula:

$$\text{Grade} = \frac{\text{Sum of scores acquired}}{\text{Maximum score}} \times 100$$

The last instrument was the assessment of product of painting. To measure the students' performance in producing works of painting, the blueprint dimensions included the description of the painting product and the character that the students gained. The dimensions were as follows: (1) the usage of art elements and principles of design, (2) creativity and originality, (3) efforts and tenacity, (4) skills, and (5) responsibility. This product assessment used self-assessment by the students. Table 4 shows the dimensions, descriptors, and an indicators of the assessment of painting products.

No.	Dimensions	Descriptors	Indicators
1.	Usage of art elements and principles of design	Painting done based on the usage of art elements and principles of design	Painting is done based on the usage of art elements and principles of design in accordance with expression of thought and feeling.
2	creativity and originality	Painting done creatively and showing originality	Painting is done through personal experimentation thoroughly in the development of forms of objects, composition, and usage of materials and tools.
3	Efforts and tenacity	Painting done with efforts and tenacity	Painting is done with maximum efforts and tenacity in development of the forms of objects, composition, and usage of materials and tools.
4	Skills	Painting done with proper skills	Painting is done based on procedures and techniques in accordance with characteristics of the materials and tools used.
5	Responsibility	Painting done with sincerity and independence	Painting is done with sincerity and independence both in the preparation and the execution.

INSTRUMENT OF SELF-ASSESSMENT OF PAINTING PRODUCT

Name of student :
 Name of School :
 Teacher (adviser) :

Name of Studio

 Owner (Painter) :
 Studio Address :

Answer the following questions by giving a check mark (✓) in the box according to your own opinion.!

1. Have you done your paintings using art elements and principles of design in accordance with expression of thought and feeling?	<input type="checkbox"/>	I have done all my paintings with awareness of the usage of art elements and principles of design in accordance with expression of thought and feeling?
	<input type="checkbox"/>	I have done most of my paintings based on awareness of the usage of art elements and principles of design in accordance with expression of thought and feeling?
	<input type="checkbox"/>	I have done some of my paintings based awareness of the usage of art elements and principles of design in accordance with expression of thought and feeling?
	<input type="checkbox"/>	I have never done my painting based on awareness of the usage of art elements and principles of design in accordance with expression of thought and feeling?

2. Have you done your paintings through personal experimentation thoroughly in the development of forms of objects, composition, and usage of materials and tools?	<input type="checkbox"/>	I have done all of my paintings through personal experimentation thoroughly in the development of forms of objects, composition, and usage of materials and tools?
	<input type="checkbox"/>	I have done most of my paintings through personal experimentation thoroughly in the development of forms of objects, composition, and usage of materials and tools?
	<input type="checkbox"/>	I have done some of my paintings through personal experimentation thoroughly in the development of forms of objects, composition, and usage of materials and tools?
	<input type="checkbox"/>	I have never done my painting through personal experimentation thoroughly in the development of forms of objects, composition, and usage of materials and tools?
3. Have you done your paintings with maximum efforts and tenacity in development of the forms of objects, composition, and usage of materials and tools?	<input type="checkbox"/>	I have done all of my paintings with maximum efforts and tenacity in development of the forms of objects, composition, and usage of materials and tools?
	<input type="checkbox"/>	I have done most of my paintings with maximum efforts and tenacity in development of the forms of objects, composition, and usage of materials and tools?
	<input type="checkbox"/>	I have done some of my paintings with maximum efforts and tenacity in development of the forms of objects, composition, and usage of materials and tools?
	<input type="checkbox"/>	I have never done my painting with maximum efforts and tenacity in development of the forms of objects, composition, and usage of materials and tools?
4. Have you done your paintings based on procedures and techniques in accordance with characteristics of the materials and tools used.	<input type="checkbox"/>	I have done all of my paintings based on procedures and techniques in accordance with characteristics of the materials and tools used.
	<input type="checkbox"/>	I have done most of my paintings based on procedures and techniques in accordance with characteristics of the materials and tools used.
	<input type="checkbox"/>	I have done some of my paintings based on procedures and techniques in accordance with characteristics of the materials and tools used.
	<input type="checkbox"/>	I have never done all of my painting based on procedures and techniques in accordance with characteristics of the materials and tools used.
5. Have you done your painting with sincerity and independence both in the preparation and the execution?	<input type="checkbox"/>	I have done all of my paintings with sincerity and independence both in the preparation and the execution?
	<input type="checkbox"/>	I have done most of my paintings with sincerity and independence both in the preparation and the execution?
	<input type="checkbox"/>	I have done some of my paintings with sincerity and independence both in the preparation and the execution?
	<input type="checkbox"/>	I have never done my painting with sincerity and independence both in the preparation and the execution?

Figure 4. Instrument of Self-Assessment Painting Product of Job Training

Hasil penilaian diri untuk menilai produk Prakerin seni lukis selanjutnya dinilai dengan formula sebagai berikut.

$$\text{Nilai} = \frac{\text{Jumlah skor yang diperoleh}}{\text{Jumlah skor maksimum}} \times 100$$

The instruments of preparation, process, and product of painting production by the students of vocational high school in the context of job training equipped with the guidelines for their usage were packed in a book which is ready to be tested further in the real setting.

IV. CONCLUSIONS AND SUGGESTIONS

A. *Conclusions*

Conclusion of this study are as follows:

1. The establishment of a set non-test instruments for assessing the painting production by the students of vocational high school which are integrated with the character assessment.
2. The establishment of guidelines for the usage of the non-test instruments for assessing the painting production by the students of vocational high school which are integrated with the character assessment in the context of job training.

B. *Suggestions:*

1. The non-test instruments for assessing the painting production by the students of vocational high school integrated with the character assessment should be tested in terms of readability by the teachers (advisers) as well as the instructors or studio owners (painters).
2. Support of administration will be needed in conducting a limited test of the instruments.

BIBLIOGRAPHY

- [1] Borg, W.R. and Gall, M.D. (1983). *Educational Research: An Introduction*. London: Longman, Inc.
- [2] Samani, Muchlas dan Hariyanto (2011). *Konsep dan Model Pendidikan Karakter*, Bandung: PT Remaja Rosda Karya.
- [3] Tranggono, Indra. (26 April 2010), Pendidikan Karakter. *Kedaulatan Rakyat*.
- [4] Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.

ANALYZING THE QUALITY OF ENGLISH TEST ITEMS OF DAILY, MID SEMESTER AND FINAL SCHOOL EXAMINATIONS IN BANDAR LAMPUNG: (ASSESSMENT AND EVALUATION IN LANGUAGE TEACHING)

Ujang Suparman
Lampung University, INDONESIA

Abstract- The problem of the current study is that the test items used to measure the objectives of the teaching and learning in the English (such as what happens in schools in Lampung Province), are rarely pre-determined before they are used, and the results of the test are also rarely analyzed systematically and professionally. Besides, few eachers of English are aware of the importance of good quality of test items such as validity, reliability, discriminating power and level of difficulty. Even some teachers consider that analyzing such matters is time consuming and hard to do. Consequently, they rarely analyze either the test items or the results of the test. The objective of the research is to potray some examples of quality of the test items using *Iteman*. The current presentation is based on a huge study using the *iteman* in English test items used in Elemntary, Junior High and Senior High Schools in Lampung. It has been found that in general the test items has high validity, shown by the Alpha in each test used for each level, Standar Deviation, average mean of each item, and Mean Biserial (average mean of the whole test items). Besides, it has also been found that some items can be used directly without any revision; some need revision before being ussed; and even some others need dropping due to the quality of the test items. It is recomended that teachers, and instructors who test their students using multiple choice items should consider the quality of test items using one of standard techniques such as the *iteman software* as an effective alternative for item test analysis because it is easy, fast, and comprehensive analysis to do.

Keywords: *item quality, option quality, validity, reliability, descriminating power*

I. INTRODUCTION

Evaluation plays a major role in education. It shows whether the objectives of the teaching and learning can be achieved or not. Therefore, the instrument used should meet the criteria of a good test item (Suparman, 2013). However, the teachers are sometimes unaware of the importance of the quality of the test they use, consequently, some of them rarely do test the instruments before they use them (see Suparman, 2013).

This research covers the analysis of huge and diverse English test items used in elemntary, junior high and senior high schools. There are 45 university English Educaion Program students participating in data collection and data analysis held in 2015 under the guidance of a lecturer of English Teaching Assessment. It involved 15 schools and 1,800 students. All the test items are used for either daily test (prepared by the English teacher), mid semester exam, final semester exam, or final school exam (UAS) usually prepared by MKKS. The objective of the research was to identify the quality of test items and to develop them based on the information obtained from the results of the analysis using the *Iteman*. The analysis covers the major issues closely related to the assessment: validity, reliability, discriminating power and level of difficulty of test items; besides it also includes the analysis of all the options comprising of the key answer to the question, and distractors. The current presentation is only a part of the total research but which can reflect and represent the figure of the quality of English test items used in Lampung.

II. THEORETICAL FOUNDATION

Many factors may influence the results of a test. For example, Kheirzadeh, et al (2015) state that the condition of a test administration, that is, the timing of the test, the testing venues and the

B. Steps of Analysis With An Iteman

The *iteman* program can work only with multiple choice items. It is relatively easy to analyze test items using the *iteman* program. The most important thing to do is to be very careful in entering the data, because if you enter the data wrongly, it would produce wrong results of data analysis. The following are the steps to enter the data using a new file:

1. Click *Start*
2. Select *Program*
3. Select *Accessories*
4. Choose and click *Notepad*
5. Save/ Click *File*
6. Select and click *Save as*, then name the data file, for example: Advread (make sure the file name must not exceed 8 letters/numbers)
7. Startt data entry, it will be faster if you work with your friend – one of you reads students' answers and the other types them. If you work with your friend, please make sure to prounuce the letter clearly, e.g., a for apple; b for ball; c for charlie; d for doctor; and e for ent.
8. It's advisable for you to save it frequently by clicking *File* and then *Save* so that the typed data will not loss if the current suddenly cuts off.
9. The data will appear like shown on the Fig. 1 above.

1. Procedure of Data Analysis Using *Iteman* Program

The steps used in the current study are as follows:

1. Open *iteman* Program, by clicking *Start*,
2. Select *Program*/click *iteman*.
3. Type the name of your data file (input) as you like on *Enter the name of the input file*. For example D:\English.txt then *Enter*.
4. Enter the name of the output file on *Enter the name of the output file*. For example, inthis case: D:\English.output then click *Enter*.
5. A question will appear, *Do you want the scores written to a file? (Y / N)*. Then type Y and click *Enter*.
6. Enter the name of your score file on *Enter the name of the score file*: For example, D:\English.scr Then click Enter. Finish. Have a good try!

The data would appear like that in the following page.

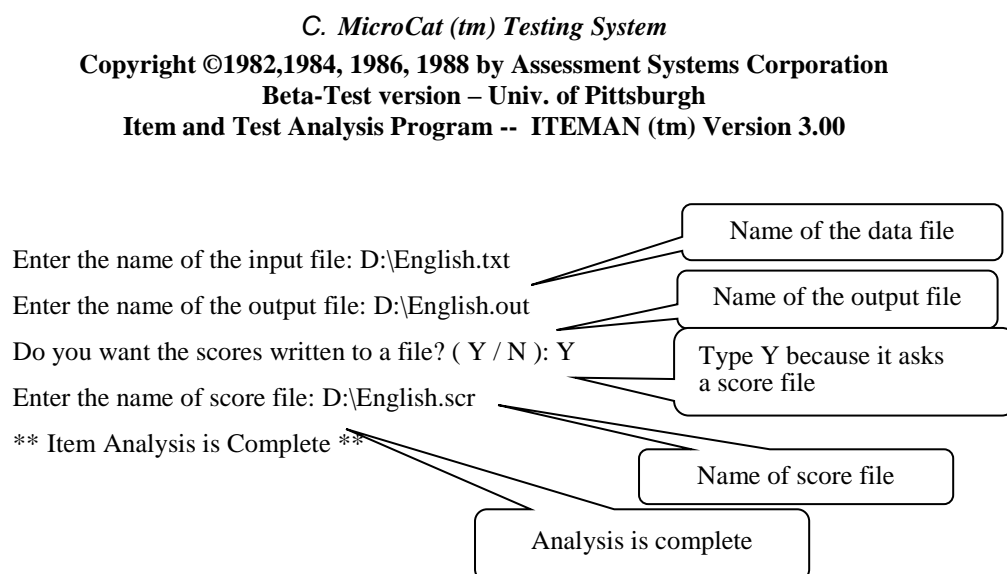


Figure 2. Item analysis appearance using ITEMAN

2. The Results of Item Analysis with Iteman

The following is the steps that the researcher used to open the results of item analysis on MS Words program:

1. Click *Start*,
2. Select *Program*/click Microsoft Word
3. Click *File*/click *Open*, please look for the results on, for example, Drive D (depends on which one you choose).
4. The following is an example of the appearance of the results of test items analysis.

MicroCAT (tm) Testing System Page 1
 Copyright (c) 1982, 1984, 1986, 1988, 1993 by Assessment Systems Corporation
 Item and Test Analysis Program -- ITEMAN (tm) Version 3.50
 Item analysis for data from file F:\Advraed.TXT
 Date: 10/22/12 Time: 8:35 am

Item Statistics					Alternative Statistics					
Seq. No.	Scale	Prop. -Item	Disc. Correct	Point Index	Point Biser.	Prop. Alt.	Endorsing Total	Point Low	Point High	Biser. Key
1	0-1	.32	.35	.36	A	.16	.30	.00	-.31	
				B	.32	.20	.55	.36	*	
				C	.03	.00	.09	.12		
				D	.49	.50	.36	-.15		
				Other	.00	.00	.00			
2	0-2	.19	.17	.11	A	.05	.00	.09	.19	?
					B	.70	.80	.55	-.27	
					C	.19	.10	.27	.11	*
					D	.05	.10	.09	.16	
					Other	.00	.00	.00		
					CHECK THE KEY					
					C was specified, A works better					
4	0-4	.27	.45	.26	A	.22	.30	.27	.10	
				B	.03	.00	.00	-.02		
				C	.27	.00	.45	.26	*	
				D	.49	.70	.27	-.31		
				Other	.00	.00	.00			
5	0-5	.78	.10	.19	A	.78	.90	1.00	.19	*
				B	.00	.00	.00			
				C	.16	.10	.00	-.17		
				D	.05	.00	.00	-.07		
				Other	.00	.00	.00			
6	0-6	.68	.41	.30	A	.08	.10	.09	.12	
				B	.05	.00	.00	-.01		
				C	.68	.50	.91	.30	*	
				D	.19	.40	.00	-.44		
				Other	.00	.00	.00			

And so on.

In the following, the resume of the results of the item tests analysis is presented and at the right side are the scores obtained by each of the participants.

Table 1. The resume of the results of the item tests analysis

Scale:	:	0	<table><thead><tr><th>No</th><th>Scores</th><th>Marks</th></tr></thead><tbody><tr><td>1</td><td>37</td><td>62</td></tr><tr><td>2</td><td>24</td><td>40</td></tr><tr><td>3</td><td>45</td><td>75</td></tr><tr><td>4</td><td>40</td><td>67</td></tr><tr><td>5</td><td>20</td><td>33</td></tr><tr><td>6</td><td>33</td><td>55</td></tr><tr><td>7</td><td>45</td><td>75</td></tr><tr><td>8</td><td>19</td><td>32</td></tr><tr><td>9</td><td>41</td><td>68</td></tr><tr><td>10</td><td>31</td><td>52</td></tr></tbody></table>			No	Scores	Marks	1	37	62	2	24	40	3	45	75	4	40	67	5	20	33	6	33	55	7	45	75	8	19	32	9	41	68	10	31	52
No	Scores	Marks																																				
1	37	62																																				
2	24	40																																				
3	45	75																																				
4	40	67																																				
5	20	33																																				
6	33	55																																				
7	45	75																																				
8	19	32																																				
9	41	68																																				
10	31	52																																				
N of Items	:	60																																				
N of Examinees	:	37																																				
Mean	:	37.703																																				
Variance	:	51.506																																				
Std. Dev.	:	7.177																																				
Skew	:	-0.677																																				
Minimum	:	19																																				
Maximum	:	51																																				
Median	:	38																																				
Alpha	:	0.798																																				
SEM	:	3.224																																				
Mean P	:	0.628																																				
Mean Item-Tot.	:	0.288																																				
Mean Biserial	:	0.405																																				
Max Score (Low)	:	34																																				
N (Low Group)	:	10																																				
Min Score (High)	:	41																																				
N (High Group)	:	11																																				

Students' scores

More importantly, ITEMAN can function as a powerful technique available to teachers for improving the quality of instruction. To achieve this, the items that would be analyzed should fulfill the following requirements: first, they have to be valid measures of instructional objectives; secondly, they have to be diagnostic, in the sense that, knowledge of which incorrect options that the students choose must be a clue to the nature of the misunderstanding, and, therefore, prescriptive of appropriate remediation; and finally, teachers who construct their own examinations may greatly improve the effectiveness of test items and the validity of test scores if they select and rewrite their items on the basis of item performance data.

D. Validity

One of the characteristics of a good test is validity. It requires a test to be usable to measure what it is intended to measure and nothing else (Power, 2012). In a similar concept, Hatch, et al (1982:250-1) define validity as "the extent to which the results of the procedure serve the uses for which they were intended." They further divide the validity into three basic types: content validity, criterion-related validity and construct validity. *Content validity* can be defined as the extent to which a test measures a representative sample of the subject matter content, that is, the test items should be relevant to the materials covered in the course. For example, in the case of the reading comprehension, the test should include the sub-skills of reading as stated in the syllabus, among others: identifying the main idea, identifying specific information relating to *who*, *what*, *when*, *why*, *where* and *how* questions; making predictions, and making inferences. The second type of validity is *criterion-related* validity. It is defined as the criteria of a test when test scores will be used to predict future performance or to estimate current performance on some valued measure other than the test itself. For example, we have designed a new language aptitude test for the students of English Study Program. And the test is thought to be a good one. Then the test is administered to a group of newly enrolled students at the English Program, and to prove that it is a valid test, the results are compared with an established test, say English Language Placement Test, which is the criterion expected to be able to predict. We predict from our aptitude test scores to performance on the major subjects at the English Program.

The two types of validity above enable us to determine how well test scores represent certain learning objectives (content validity) or how well they predict or estimate a certain performance (criterion-related validity). Besides these more specific and practical uses, sometimes the validity of 'certain general psychological construct' (Hatch, et al, 1982: 252) needs to be identified. For example, when the students' performances in terms of psychological aspects (such as *self esteem*, *extrovert/introvert*, *acculturated*, *motivated*) need to be interpreted, and how important they are in language learning in English classes), then construct validity is required. But this type of validity is not the concern of the current study.

E. Reliability

A clear cut and direct definition of reliability is "consistency of measurement" (Su, et al, 2015), In this study the definition of *reliability* is straightforward: a measurement is reliable if it represents mostly true score, relative to the error. For example, an item such as "Red foreign cars are particularly ugly" would likely provide an unreliable measurement of prejudices against foreign-made cars. This is because there probably are many individual differences concerning the likes and dislikes of colors. Thus, this item would "capture" not only a person's prejudice but also his or her color preference. Therefore, the proportion of true score (for prejudice) in subjects' response to that item would be relatively small.

At least, *Reliability & Item Analysis* have three major functions. First, they may be used to construct reliable measurement scales, secondly, to improve existing scales, and finally to evaluate the reliability of scales already in use. In a more specific objective, *Reliability & Item Analysis* will aid in the design and evaluation of *sum scales*, that is, scales that are made up of multiple individual measurements (e.g., different items, repeated measurements, and different measurement devices). Numerous statistics can be computed to allow us to build and evaluate scales following the so-called *classical testing theory* model.

F. Measures of Reliability

Based on the discussion above, one can easily infer a measure or statistic to describe the reliability of an item or scale. Specifically, an *index of reliability* may be defined in terms of the proportion of true score variability that is captured across subjects or test takers, relative to the total observed variability. In equation form, it can be stated: Reliability = $\frac{\sigma^2_{\text{(true score)}}}{\sigma^2_{\text{(total observed)}}}$. but this equation is not used in this study because that is automatically calculated by the *iteman* software.

G. Number of Items and Reliability

This concept describes a basic principle of test design. That is, the more items there are in a scale designed to measure a particular concept, the more reliable will the measurement be. Let us examine the following example to clarify the concept. Suppose you want to measure the height of 10 persons, using only a crude stick as the measurement device. Note that we are not interested in this example in the absolute correctness of measurement (i.e., in inches or centimeters), but rather in the ability to distinguish reliably between the 10 individuals in terms of their height. If each person is measured only once in terms of multiples of lengths of your crude measurement stick, the resultant measurement may not be very reliable. However, if each person is measured 100 times, and then take the average of those 100 measurements as the summary of the respective person's height, then you will be able to make very precise and reliable distinctions between people (based solely on the crude measurement stick).

H. Discriminating Power

There are two indicators of the item's discrimination effectiveness: point biserial correlation and biserial correlation coefficient (Matlock-Hetze, 1997). The choice of correlation is determined by what kind of question we want to answer. The advantage of using discrimination coefficients over the discrimination index (D) is that every person taking the test is used to compute the discrimination coefficients and only 54% (27% upper + 27% lower) are used to compute the discrimination index, \underline{D} .

The point biserial (r_{pbis}) correlation is used to find out if the right people are getting the items right, and how much predictive power the item has and how it would contribute to predictions. The discriminating power (D) of test items can be measured by one of the three ways: discriminating index; correlation index; and harmonious index. A discriminating power is usually symbolized with a capital D, which can be determined by the following steps: First, rank order the answer sheet top-down from the highest to the lowest scores based on the total number of test takers; then multiply N with 27%, the results is n score; after that, calculate n from the Upper Group (the answer sheets with high scores are counted from the top) while n from the Lower Group (the answer sheets with low scores are counted from the bottom). And finally, determine the proportion of the test items answered correctly by each group. That is, the correct answers from each of the Upper Group (pU) and Lower Group (pL) are divided by n. the discriminating power is in fact the differences of the proportion of the correct answers between the UG and the LG. So, it can be stated that $D = pU - pL$.

To determine whether a test item is accepted, revised or rejected, the following parametric criteria is used:

**Parameter of Decision
D Coefficient**

$D = > 0.30$	accepted
$D = 0.10 - 0.29$	revised
$D = < 0.10$	rejected

I. Level of Difficulty

Level of item difficulty can be defined as the percentage of students taking the test who answered the item correctly. In short, it can be stated that the larger the percentage getting an item right, the easier the item. The higher the difficulty index, the easier the item is understood to be. Matlock-Hetzel (1997) states that to compute the item difficulty, the examiner can divide the number of people answering the item correctly by the total number of people answering item. The proportion for the item is usually denoted as p and is called item difficulty (Crocker & Algina, 1986). An item answered correctly by 85% of the examinees would have an item difficulty, or p value, of .85, whereas an item answered correctly by 50% of the examinees would have a lower item difficulty, or p value, of .50.

The easiest way to measure the level of difficulty of an item is by using proportional scale or proportion correct (p), that is, the number of test takers answering correctly on the items under analysis is compared with the total number of test takers. The equation is as follows:

$$p = \frac{\sum B}{N}$$

where p = the proportion of test takers who answer correctly a certain item under analysis

$\sum B$ = the number of test takers who answer correctly

N = the total number of test takers.

The level of difficulty ranges from 0 through 1. It can be categorized into three classifications as follows:

Proportion Correct (p)	Category
$p \geq 0.70$: easy
$0.30 < p < 0.70$: Average
$p < 0.30$: difficult

III. METHOD

The design of the research is descriptive assessment, that is, a study describing the results of an analysis of the topic under discussion, which was adjusted with standardized criteria. The research analyzed the test items used to assess the students' ability to response to daily, mid semester, final semester and final school examinations in elementary, junior high, and senior high schools. The tests were prepared by the teachers of English for daily, mid semester and final semester examinations. Whereas the tests for final school examination (UAS) are usually prepared by the board of headmasters (MKKS). The research took place in the schools mostly in Bandar Lampung; some schools are out of Bandar Lampung. It was carried out during the First Semester of the 2015/2016 academic year. The researcher collaborates with 45 students from the English Study Program, University of Lampung taking English Teaching Assessment course. The total participants of the research consist of 1,800 students learning English classes in 15 schools, each level (elementary, junior high and senior high schools).

The data were collected by means of documentation, that is, using the students' answer sheets on the English tests or examination comprising daily English tests, mid semester English test, final semester examination and final school examination on the academic year mentioned above. The data, that is, the students' answers and scores on English tests were analyzed using

Iteman software. The analysis covered four major issues relating to the assessment: validity, reliability, discriminating power and level of difficulty.

A. How to Interpret the Results of Item Analysis

Based on the recommendations from some experts of measurement, the following criteria to determine the quality of test items and its interpretation have been agreed:

to classify which test items can be used directly without prior revision, which ones need revising or even dropping.

Table 2. Criteria to classify the quality of test items

Level of Difficulty (p)	
0.000 – 0.099	Very difficult/needs total revising
0.100 – 0.299	Difficult/needs revising
0.300 – 0.700	Average/good
0.701 – 0.900	Easy/needs revising
0.901 – 1.000	Very easy/ needs dropping or total revising
Point Biserial (Discriminating Power – D)	
0.199 –	Very low $\leq D$ /needs dropping or total revising
0.200 – 0.299	Low/needs revising
0.300 – 0.399	Quite average/without revision
0.400	High $\geq D$ /very good
Prop Endorsing (proportion of the answers)	
0.000 – 0.010	Least/drop, or needs revising
0.011 – 0.050	Sufficient/good enough
0.051 – 1.000	Very Good
Alpha (reliability of test item)	
0.000 – 0.400	Low/not sufficient
0.401 – 0.700	Average/sufficient
0.701 – 1.000	High/Good

IV. FINDINGS AND CONCLUSIONS

Based on the results of the data analysis on daily test, mid semester, final semester, and final school examinations, the following are found:

A. Example of Daily Test Results of Analysis

In one of elementary school, the results of test items analysis show that out of 30 English test items 8 items (good, not necessarily revised before being used), 9 items should be revised because the quality is not sufficient for a good test and 13 items should be dropped because the quality is too bad. The following figure shows the quality of the daily test items.

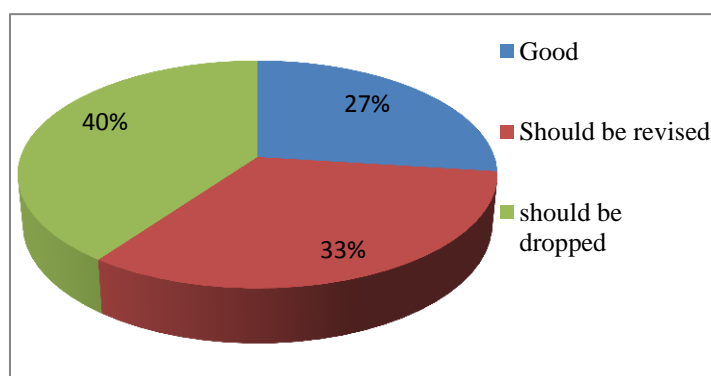


Figure 3. Quality of Daily Test Items in an Elementary School in Lampung

The figure above shows that good test items are relatively smaller in number than the items that should be revised. And what makes things worse is that the items that should be dropped have the highest percentage. It is a challenge for teachers, school supervisors and Diknas to solve the problem.

B. Example of Mid Semester Exam Results of Analysis

In one of the state junior high schools, it was found that in a mid semester exam, there were 17 out of 50 test items (34%) which can be used directly without any revision.

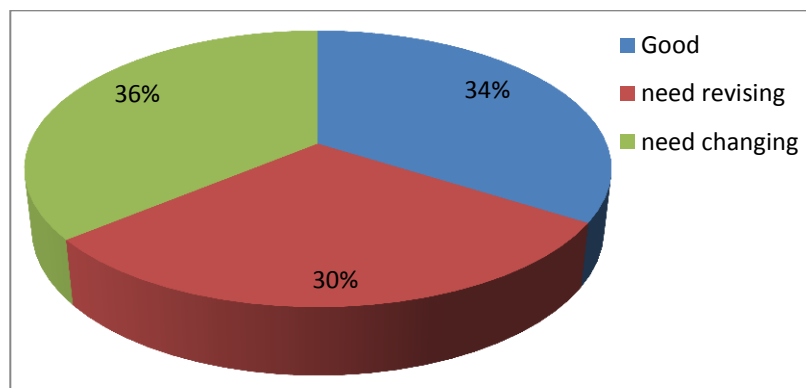


Figure 4. Quality of Mid Semester Exam Items in an Elementary School in Lampung

Besides, there are 15 out 50 English test items (30%) which should be revised before they are used in the examination. And unexpectedly, there are 18 out of 50 test items (36%) which should be dropped because their quality were very bad. There are several reasons for the problem including: The range between Prop Correct and Point Biserial is too far; the values of Point Biserial are undetected by the iteman because it is too small; there are negative points (minus) in several Point Biserial values; the values of Prop Correct is very high, even can reach 1.00. It means that the item is very easy. Both of the values of Prop Correct and Point Biserial are very low. There are some 0 values found in Point Biserial, which means that there is no different range in answering/choosing option. Low students and hig students all choose the right answer.

C. Example of Final Semester Exam Results of Analysis

The results of the data analysis in a final semester examination in one of senior high schools show that 6 out of 50 test items (12%) are good and therefore can be used directly without any revision; 17 out of 50 (34%) need revising because they are not good enough based on the predetermined criteria; and 27 out 50 (54%) should be dropped because the quality of the items are very bad. This is represented by the following figure.

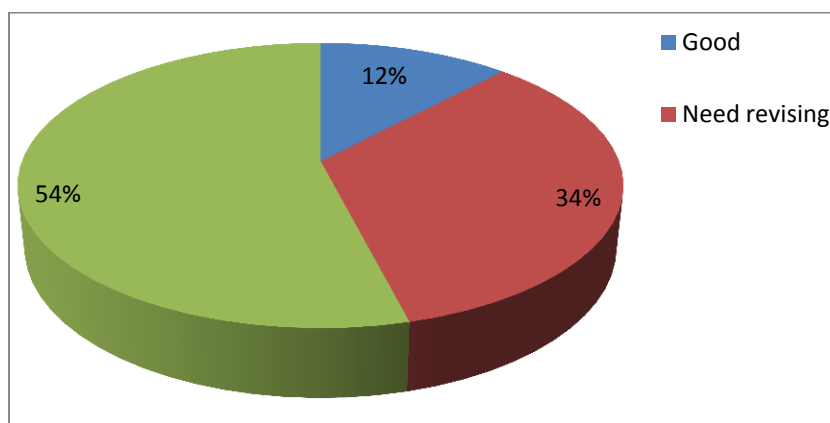


Figure 5. Quality of Final Semester Exam Test Items in a Senior High School School in Lampung

The figure also shows that unexpectedly 54% of the total number of the test items should be dropped which means that majority of the English test items used in a final semester examination are not good.

D. Example of Final School Exam Results of Analysis

Another result of data analysis of English test items used in a final school examination (UAS) has also shown similar findings to what has been discussed so far as illustrated by the following figure.

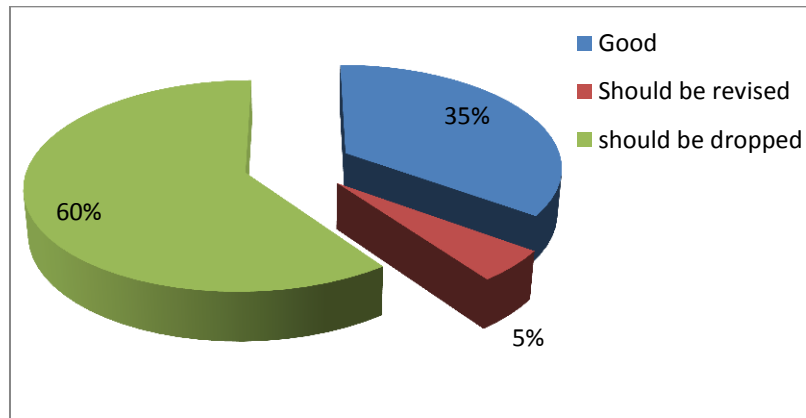


Figure 6. Quality of Final School Examination Test Items in a Senior High School School in Lampung

Like what has been explained so far, the last figure above also shows that majority of the English test items (60%) should be dropped because the quality of them are very bad.

Based on the findings above the following conclusions can be drawn and recommendations can be put forward:

1. Although the analysis of the English test items was carried in different levels (elementary, junior high and senior high schools), the main stream is similar, that is, the quality of English test items absolutely needs improving.
2. In all levels of educational institutions, the English test items can be categorized into three classifications: sound, that is, they can be used directly without prior revision, unfortunately the percentage is relatively low; need revising before being used (unexpectedly the percentage is very high); and need dropping because they are too bad (unfortunately the percentage is sometimes very high).
3. Given the teachers are very busy, they seem hardly try out the test items that they will use. They tend to design the tests and directly use them.
4. There is no enforcement from the school supervisors and/or Diknas for teachers to try out the test items, analyze the results and make sure the quality of the items.
5. Although evaluation belongs to one of the 8 standards of national education, in reality its implementation in the lowest layer – school – seems, to some extent, neglected.
6. It is recommended that teachers of English in all levels of educational institutions should always try out the tests that they have prepared before being used to measure the objectives of their teaching.
7. There should be special training for English teachers on how to analyze test items so that they can make sure that the tests that they design are sound – having high quality based on the standard of a good test, and whether their teaching objectives can be achieved.
8. The school supervisors and/or Diknas should put priority on the improving the quality of test items so that the quality of teaching and learning process can be improved.
9. School headmasters should make sure regularly at least once in one semester that the teachers have done the try out for any kind of test that they will use.
10. To improve English teachers' motivation, there should be working appraisal for those who have done the tryout of the tests and analyzed the results.

BIBLIOGRAPHY

- [1] ASC. 1989-2006. *User's Manual for the ITEMAN™ Conventional Item Analysis Program*. St. Paul, Minnesota: Assessment Systems Corporation
- [2] Crocker, L., and Algina, J. 1986. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- [3] Hatch, E. & Farhady, H. 1982. *Research design and statistics for applied linguistics*. Rowley, Massachusetts: Newbury House Publishers.
- [4] Kheirzadeh, S., Marandi, S.S., & Tavakoli, M. 2015. Test Administration Conditions of the General English Section of the Iranian National PhD Entrance Exam: Are the PhD Exam Candidates Satisfied?. *Journal of Language Testing* Vol. 5, No. 2, October 2015
- [5] Lyman, H.B. 1971. *Test scores and what they mean*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.
- [6] Matlock-Hetzel, S. 1997. *Basic Concepts in Item and Test Analysis*. Texas: A&M University
- [7] Miriam, K.L. 1996. *An introduction to psychological tests and scales*. London: UCL Press.
- [8] Ngadimun, H. 2004. *Analisis butir soal dengan komputer dan menafsirkannya*. Makalah disampaikan pada sosialisasi KBK bagi guru SMP Kabupaten Tanggamus di Pulau Panggung, tanggal 22-24 Juli 2004. Bandar Lampung: HEPI.
- [9] Power, T. 2012. *Methods of assessment*. <http://www.tedpower.co.uk/esl0706.html> (Accessed: 09 May 2013)
- [10] Su, Y & Shin, S.Y. 2015. Test Review: The New HSK. *Iranian Journal of Language Testing*. Vol. 5, No. 2, October 2015
- [11] Suparman, U. 2011. The implementation of Iteman to improve the quality of English test items as a foreign language. *Aksara*, Vol. XII, No. 1 pp. 85-96. Bandar Lampung, April 2011.
- [12] Suparman, U. 2013. *Improving the quality of English test items as a foreign language by means of Iteman: (an assessment analysis)*. A paper presented on the First International Teacher Education Conference (ITEC) organized by Lampung Univeristy in Bukit Randu Hotel, June 30th – July 3rd, 2013.

Developing A Pedagogical Commitment Instrument

Wasidi

Faculty of Educational and Teacher Training - Bengkulu University
wasidirma@gmail.com, wasidi@unib.ac.id

Abstract. The purpose of this study was to develop the valid and reliable pedagogical commitment instrument. The benefit of this development was availability of pedagogical commitment instrument that can be used to educational institutions. This research followed the Gable method. The construct of pedagogical commitment instrument was validated by seven experts. Trial phase carried out on 630 teacher students on Bengkulu University. Data were analyzed by using content validity, CFA, IRT PCM, concurrent validity, convergent validity, and reliability coefficient. The result showed that the pedagogical commitment instrument has content validity index of 0.842. The instrument has a goodness of fit statistics, and pedagogical commitment instrument eligible partial credit model. The concurrent validity of 0.355, the convergent validity of 0.842, and the reliability coefficient of 0.701. Thus pedagogical commitment instrument has qualified content validity, construct validity, concurrent validity, and convergent validity, and reliability coefficient.

Keywords: *pedagogical commitment, content validity, CFA, concurrent validity, convergent validity, reliability coefficient.*

I. INTRODUCTION

Implementation of the Law number 20 in 2003 about educational system, law number 14 in 2014, government regulation number 19 in 2005 requires that teachers have the ability ideal among other aptitudes, interests, calls the soul, and idealism [1]. Sudarnoto mentions that the teacher is a crucial factor in the success of education as assessed student achieve [2]. The result of research by Block showed that mentions the role of the teacher is very great in the changing world a growing body of research indicates that organizational commitment has utility as a predictor of such important behavioral outcomes as performance [3].

Teacher commitment is very vital in the performance of mathematics. Teachers who demonstrate high level of commitment to their profession teach effectively, thus bringing about good performances amongst their students [4]. An important finding of our research is that the teachers' commitment is positively correlated with the wellbeing and achievement of children with disabilities. Other reasons for teachers' commitment are the positive emotional attachment to children with disabilities and their families and the responsibility to maintain professional knowledge [5]. These results indicate that teachers' commitment from special schools is a predictor or an echo of their humanity and loyalty. The results show that the relationship between instructional leadership and the level of teachers' commitment is high. There is a significant relationship between instructional leadership and the level of teachers' commitment in three MJSCs in Pahang. In terms of the implications of the study, the principal should use his own intelligence in instructional leadership skills to develop teachers' commitment [6]. Pedagogical commitment effect on student performance, but in Indonesia is not available pedagogical commitment instrument yet, it is necessary to develop the instruments.

Task commitment represents energy brought to bear on a particular problem (task) or specific performance area [7]. Committed teachers may have strong psychological ties to their school, their students, or their subject areas. In loosely coupled organizations like schools, where there is considerable disagreement on what outcomes should be accomplished and great difficulty in inspecting and controlling work [8].

Teacher commitment consist of the three dimensions in the empirical study: commitment to school mission, to school-community partnerships, and to the school as a learning community [9]. Four factors influencing teachers' commitment were found, viz. "students' learning and school development", "demands on teaching and school practices", "teaching as a career choice", teacher-pupil interaction and attitudes [10]. Binding themselves to the task of a recognition of the

task as a form of internal motivation that drives a person to be diligent and tenacious in their work, despite the many obstacles [11].

The image of a committed teacher was one who gives extra time, thought and effort for students, school, and teaching; is willing to seek improvement, to do things to an excellent standard; is engaged whole-heartedly, more than merely to fulfill responsibility; has a vision and strives to do something right for the students [12]. A committed teacher was likely to: be less tardy, work harder, and be less inclined to leave the workplace; devote more time to extra-curricular activities in order to accomplish the goals of the educational organization; perform work better; influence student achievement; believe and act upon the goals of the school; exert efforts beyond personal interest; and intend to remain a member of the school system [13]. A teacher who is committed to students and makes efforts to create a supportive learning climate in the classroom, for example, may not have affective affiliation to the school organization due to the lack of identification with school goals. Thus a teacher highly dedicated to student affairs but resistant to school reforms may be regarded as committed when her/his attitude towards students is assessed. She/he may not be considered a committed teacher if the assessment is taken from the perspective of school organization [14].

Teachers may affect the overall level of social preference in their classroom because elementary school children's evaluations of their peers may be based, in part, on their observations of the teacher's reactions to these students [15]. Education researchers list some characteristics of committed teachers and principals for their organizations as teacher commitment: teachers constantly should talk about developing teaching approaches, teachers and principals should continuously observe each other's performance, and evaluate their own performances through these observations. Teachers and principals should together supply education materials, teachers and principals should help each other with development of teaching approaches [16].

Frequent teacher-family communication immediately increased student engagement as measured by homework completion rates, on-task behavior, and class participation [17]. Teachers' commitment and enthusiasm are qualities that are harder to measure but are seen to be more directly related to the quality of teaching and learning than the traditional emphases on qualifications and years of experience [18]. Schools can and should function as disciplined spaces that nonetheless reflect the realities of youths' lived experience, as well as the aspirations of communities of color for freedom from racism, sexism, and classism [19]. In schools, teachers must be patient, able to motivate students, understand the students' needs, and accept differences in others. In schools, teachers are empowering their capacity for growth and success. When teachers are empowered in areas important to them, they at the same time build the capacity of students. Hence, teachers require professional skills. Communication, cooperation and commitment are essential skills [20].

II. METHOD

The pedagogical commitment instrument was developed refers to Gable method. The method consist of 15 phases, there are conceptual definition developing, operational definition formulas, scaling method, conceptual review, instrument formatting, determine sampling, first trial, first trial analysis, instrument revision, last trial, instrument finalizing, criteria related validity, manual instrument [21].

The pedagogical commitment construct consist of four factors. The each factor was translated to eight items, thus the instrument consist of 32 items. The instrument item of emotional intelligence was composed case that can be seen daily. The each item of instrument was followed by three options with gradation scoring. The each statement was followed by three possible answers graded by random to avoid choosing origin [22,23]. The draft instrument that has been validated by seven experts, there were one linguist, two psychologists, two education experts, and two educational measurement experts. The analysis used the content validity formula [24,25]. Instrument legibility trials conducted to 3 undergraduate students of Mathematics, Physics, and Culinary Art at Yogyakarta State University. The participants were teacher student of Bengkulu University. The number of participant that can be used as samples are 630 people representative all of student undergraduate. The construct of teacher aptitude instrument was analyzed by confirmatory factor analysis [26,27] and using LISREL software [28]. Item analysis of instruments used IRT PCM formula [29], and using the QUEST software [30]. The external validity using Edward Personal Preference Schedule instruments. The correlation

between emotional intelligence score and EPPS score be compute by product moment formula, and the convergent validity using mulitrait multimethod matrix [31].

III. RESULT AND DISCUSSION

The constructs of pedagogical commitment are task motivation, responsibility, discipline, and persistent. Instrument was designed by Guttman model. The each factor of pedagogical commitment be developed became 8 items statement. Statement consist of cases was followed 4 ordinal option. The item statement is followed by four graded options, with a score of 1, 2, 3, and 4. However, after obtaining input-critic and expert advice every item changed to a statement followed by three option graded with a score of 1, 2, and 3 , but the option is not sorted in order of score. The sample item instrument before and after validation is presented below.

You find students who have difficulty in understanding the lesson. How do you respond to these students?

1. It does not matter, because it is the responsibility of parents.
2. Feeling pity, but it is not my responsibility.
3. Want to help, if asked parents / teachers.
4. As a teacher called for help.

This item is improved became:

You find children very intelligent, often asking questions difficult to answer. How do you anticipate the questions the child?

1. Update knowledge of material / teaching materials
2. Invite other children to answer.
3. Questions are answered could.

The result of construct validation by expert can be displayed follow.

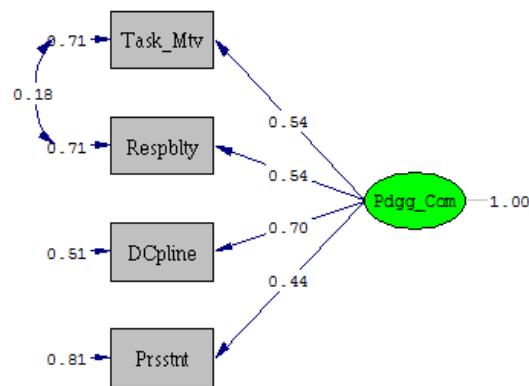
Table 1. The Content Validity Index of construct

Variable	Factors	I V	p
Pedagogical Commitment	Task Motivation	0.893	0.003
	Responsibility	0.893	0.003
	Discipline	0.821	0.012
	Persistent	0.893	0.003

Base on Table 1, all of factors of pedagogical commitment were good content validity, there were lager than 0.800 and the probability smaller than 0.05.

Table 2. The Content Validity Index of Items

Item No.	iV	p	Item No.	iV	p
1	0.857	0.006	17	0.857	0.006
2	0.821	0.012	18	0.857	0.006
3	0.857	0.006	19	0.893	0.003
4	0.821	0.012	20	0.857	0.006
5	0.821	0.012	21	0.857	0.006
6	0.857	0.006	22	0.821	0.012
7	0.821	0.012	23	0.857	0.006
8	0.786	0.023	24	0.893	0.003
9	0.750	0.041	25	0.857	0.006
10	0.821	0.012	26	0.786	0.023
11	0.857	0.006	27	0.821	0.012
12	0.857	0.006	28	0.821	0.012
13	0.857	0.006	29	0.821	0.012
14	0.821	0.012	30	0.893	0.003
15	0.821	0.012	31	0.893	0.003
16	0.821	0.012	32	0.893	0.003
				0.842	0.008



Chi-Square=0.19, df=1, P-value=0.65987, RMSEA=0.000

Figure 1. CFA Standardized of Pedagogical Commitment

Fig. 1 showed that loading factor of task motivation is 0.54 and t values is 9.66. The loading factor responsibility is 0.54 and t value is 9.58. The loading factor of discipline is 0.70 and t value is 11.93. The persistent loading factor is 0.44 and t value is 8.9. value of Chi-Square is 0.19 smaller than 2df, p value 0.65987 smaller than 0.05, and RMSEA is 0.00 smaller 0.08, that is means the construct of pedagogical commitment has goodness of statistics [28,32]. All of loading factor larger than 0.30, and the t value larger than 1.96, that means is valid as construct of pedagogical commitment. [33]. The correlation between task motivation and responsibility is 0.18, p value is 0.36, although both correlated variables are not great, but it shows that people are motivated tasks will have a sense of responsibility.

Base on Fig. 2, all of item has infit MNSQ value from 0.77 up to 1.18. The infit value in rate from 0.77 up to 1.3 that means all of item eligible with Rasch partial credit model.

Pedagogical Commitment 32 ITEM						
Item Fit				1/ 1/15 12:42		
all on all : N = 630 L = 32 Probability Level= .50						
INFIT						
MNSQ	.63	.71	.83	1.00	1.20	1.40 1.60
<hr/>						
1 item 1			.		*	.
2 item 2		.			*	.
3 item 3		.	*			.
4 item 4		.			*	.
5 item 5		.	*			.
6 item 6		.	*			.
7 item 7		.			*	.
8 item 8		.			*	.
9 item 9		.				.
10 item 10		.			*	.
11 item 11		.			*	.
12 item 12		.	*			.
13 item 13		.	*			.
14 item 14		.			*	.
15 item 15		.	*			.
16 item 16		.			*	.
17 item 17		.	*			.
18 item 18		.	*			.
19 item 19		*				.
20 item 20		.			*	.
21 item 21		.			*	.
22 item 22		.			*	.
23 item 23		.			*	.
24 item 24		.			*	.
25 item 25		.	*			.
26 item 26		.			*	.
27 item 27		.			*	.
28 item 28		.	*			.
29 item 29		.			*	.
30 item 30		.			*	.
31 item 31		.			*	.
32 item 32		.			*	.

Figure 2. Plot of Item Infit

Table 3. Infit and Outfit, and Standard Deviation of Item

No	Statistics	Item Estimasi	Examinee Estimasi
1	Mean	0,00	1,05
2	Standard deviation	0,77	0,49
3	Standard deviation ajusted	0,74	0,39
4	Reliability coefficient	0,92	0,71
5	Infit mean square & standard deviation	0,99 ± 0,08	0,99 ± 0,03
6	Outfit mean square & standard deviation	0,98 ± 0,07	0,98 ± 0,05
7	Infit t	0,15 ± 1,05	0,00 ± 1,08
8	Outfit t	0,08 ± 1,07	0,04 ± 0,95

Table 3 shows that the estimated grain mean square infit the mean SD of 0.99 and 0.08, while the mean square outfit gained 0.98 mean and SD of 0.07. Infit t with a mean of 0.15 and SD of 1.05, while the outfit t with a mean of 0.08 and SD 0.17. The examine estimates show that the infit mean square standard deviation of 0.99 and 0.03. The outfit square means of 0.98 and standard deviation of 0.05. The infit t of 0.00 and a standard deviation of 1.08. The outfit t of 0.04 and a standard deviation of 0.95. This means that the model is fit the Rasch model. The reliability pedagogical commitment coefficient of 0.71 larger than minimal require of 0.70. Thus whole grains pedagogical commitment eligible polytomous Rasch models, and they are eligible to be used as a good instrument.

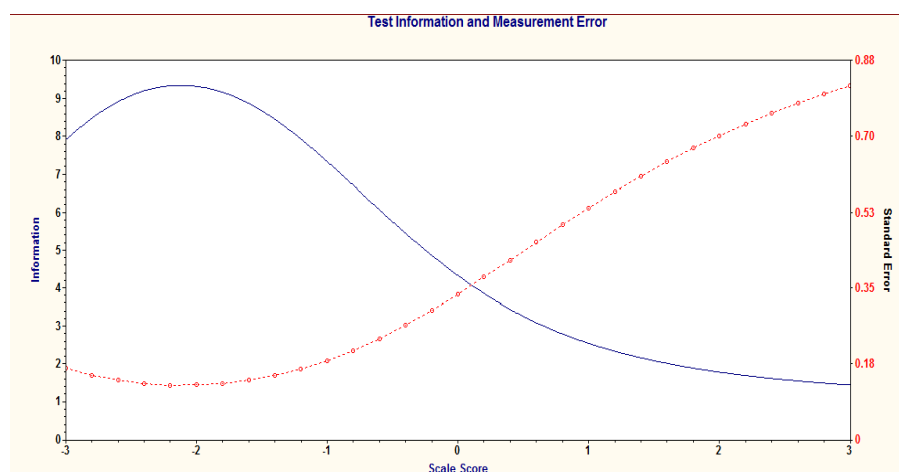


Figure 3. Total information function

The information maximum and the minimum error of measurement at a scale point -2.2. The intersection of information total lines and standard error of measurement lines at a scale score point is 0.2. Total information function was 0.2, that means the pedagogical commitment instrument eligible for ability person up to 0.2.

Table 4. The Concurrent Validity

Variabel	EPPS
Pedagogical Commitment	0,355
p	0,050*

Base on Table 4, the correlation between the pedagogical commitment score by EPPS score of 0.355. The relationship between pedagogical commitment score with an EPPS score significant at 95% confidence level. This is shown with probability one that appears at 0,050 p. The significant relationship between pedagogical commitment by an EPPS, it's mean that both of measuring the object is similar.

Base on Table 5, on the motivation factor, the correlations between tasks inventory scores and observation score of 0.679. On he task completion factor there is a correlation between the score of the inventory with the score observations lecturer of 0.713. At the factors liable to duty there is a correlation between the inventory score with a observations score of 0.740. On oomph factor - the completion of the task there is a correlation between the inventory score with a observations score of 0.736.

Table 5. The Convergent Validity

		Inventory					Observation				
		F1	F2	F3	F4	TI	F1	F2	F3	F4	TP
Inventory	F1	1	0.611	0.179	0.472	0.744	0.679	0.236	0.095	0.489	0.584
	F2		1	0.198	0.398	0.815	0.318	0.713	0.308	0.531	0.823
	F3			1	0.131	0.589	0.118	0.023	0.740	-0.058	0.435
	F4				1	0.781	0.296	0.265	0.091	0.736	0.640
	TI					1	0.611	0.563	0.556	0.682	0.842
Observation	F1						1	0.13	0.144	0.407	0.726
	F2							1	0.243	0.327	0.748
	F3								1	-0.045	0.672
	F4									1	0.742
	TP										1

At factor 1 and factor 2 motivation to the task and task completion correlation is high at 0.611, this shows that for a job well done takes a good motivation anyway, or because it can be a job well done, the motivation of the completion of the task will be high, so the relationship between the two factors was high, which means these factors measure the same properties. All correlations between factors inventory scores by factors including high observation scores. Overall correlation between inventory scores and a observations score of 0.842. This means that both of these measurements to measure the properties of the same trait.

IV. CONCLUSION

The emotional intelligence construct was eligible goodness of fit statistics. The all of instrument item qualified IRT PCM, with a reliability coefficient of 0.71 and the value of information to +0.2. The concurrent validity in enough categories, while the convergent validity of 0.842. Thus the instrument of emotional intelligence has good criteria for teacher student.

REFERENCE

- [1] Anonim, Undang-undang no 20, in: P. Nasional (Ed.), Departemen Pendidikan Nasional, Jakarta, 2003.
- [2] L.F.N. Sudarnoto, Memberdayakan profesi pendidik untuk mencapai peningkatan kualitas kehidupan kerja, Universitas Atma Jaya, Yogyakarta, 2009.
- [3] J.H. Morris, & J. Daniel Sherman, Generalizability of an Organizational Commitment Model, The Academy of Management Journal 24 (1981) 512-526.
- [4] P. Kiplagat, Elizabeth Role, & Lazarus Ndiku Makewa Teacher commitment and mathematics performance in primary schools: A meeting point, International Journal of Development and Sustainability 1 (2012) 286-304.
- [5] A. Margaritoui, Teachers' Commitment from Special-Need Schools – a Predictor of their Humanity and Loyalty, Procedia - Social and Behavioral Sciences 203 (2015) 322 – 326.
- [6] N.M. Mat Ali, Mohd Radzi bin Taib, Hazlina bt Jaafar, Wan Aida, Rohana Mohamed Salleh, Mohd Nawawi bin Omar, Principals' Instructional Leadership And Teachers' Commitment In Three Mara Junior Science Colleges (Mjisc) In Pahang, Malaysia, Procedia - Social and Behavioral Sciences 191 (2015) 1848 – 1853.
- [7] J.S. Renzulli, What Makes Giftedness?: Reexamining a Definition, Phi Delta Kappan 60 (1978) 180-184.
- [8] W.A. Firestone, & James R. Pennell, Teacher commitment, working conditions, and differential incentive policies, Review of Educational Research 63 (1993) 489-525.
- [9] J.A. Ross, & Peter Gray, Transformational leadership and teacher commitment to organizational values: The mediating effects of collective teacher efficacy, School Effectiveness and School Improvement 7 (2006) 179 – 199.
- [10] C. Kwok-wai, In-service teachers' motives and commitment in teaching, Hong Kong Teachers' Centre Journal 5 (2006) 112-128.
- [11] S.C.U. Munandar, Kreativitas dan keberbakatan : Strategi mewujudkan potensi kreatif dan bakat, Gramedia Pustaka Utama., Jakarta, 1999.
- [12] C. Pik Lin, & Sylvia Yee Fan Tang, Teacher commitment trends: Cases of Hong Kong teachers from 1997 to 2007, Teaching and Teacher Education 25 (2009) 767-777.
- [13] N. Abd Razak, I. Gusti Ngurah Darmawan, John P. Keeves, The influence of culture on teacher commitment, Soc Psychol Educ. 13 (2010) 185-205.
- [14] C. Pik Lin, & Gable, Robert K., Satisfied and dissatisfied commitment: Teachers in three generations, Australian Journal of Teacher Education 36 (2011) 43-75.
- [15] A.Y. Mikamia, Marissa Swaim Griggsa, Meg M. Reulanda, & Anne Gregoryb, Teacher practices as predictors of children's classroom social preference Journal of School Psychology 50 (2012) 95-111.
- [16] Ç.T. Mart, A passionate teacher: Teacher commitment and dedication to student learning, International Journal of Academic Research in Progressive Education and Development 2 (2013) 437-442.
- [17] M.A. Kraft, & Shaun M. Dougherty, The effect of teacher-family communication on student engagement: Evidence from a randomized field experiment, Journal of Research on Educational Effectiveness 6 (2013) 199-222.

-
- [18] G. Qing, & Christopher Day, Challenges to teacher resilience: conditions count, *British Educational Research Journal* 39 (2013) 22-44.
 - [19] K.W. Yang, Discipline or punish? Some suggestions for school policy and teacher practice, *Language Arts* 87 (2009) 49-61.
 - [20] R. Shagholia, Mohammad Reza Zabihi, Masood Atefi, & Farzaneh Moayedi, The consequences of organizational commitment in education, *Procedia Social and Behavioral Sciences* 15 (2011) 246-250.
 - [21] R.K. Gable, *Instrument Development in Affective Domain*, Kluwer-Nijhoff Publishing, Boston, 1986.
 - [22] P.D. Rankin, Gerald A. Knezek, Susan Wallace, & Shuqiang Zhang, *Scaling method*, Lawrence Erlbaum Associates Publisher, New Jersey, 2004.
 - [23] B. Sumintono, dan Wahyu Widhiarso, *Aplikasi Model Rasch untuk penelitian Ilmu-ilmu Sosial*, Tim Komunikata Publishing House, Bandung, 2013.
 - [24] L.R. Aiken, Content validity and reliability of single items or questionnaires, *Educational and Psychological Measurement*. 40 (1980) 955-959.
 - [25] L.R. Aiken, Three coefficient for analyzing the reliability and validity of rating, *Educational and Psychological Measurement*. 45 (1985) 131-142.
 - [26] R.J. Shavelon, & George C. Stanton Construct validation : methodology and application to three measures of cognitive struktur, *Journal Educational Measurement* 12 (1975) 67- 85.
 - [27] I.S. Kirsch , & John T. Guthrie, Construct validity of functional reading tests, *Journal of Educational Measurement* 1 (1980) 281-293.
 - [28] K. Joreskog, & Sorbom, D., *Lisrel 8 : User's reference guide*, Scientific software international, Chicago, 2004.
 - [29] D. Andrich, A rating formulation for ordered response categories, *Psichometrika* 43 (1978) 561-573.
 - [30] R.J. Adams, & Khoo Siek Toon, *QUEST. The Interactive Test Analysis System*, The Australian Council for Educational Research, Camberwell, Victoria, 1996.
 - [31] D.T. Campbell, & Donald W. Fiske, Convergent and discriminant validation by the multitrait-multimethod matrix, *Psychological Bulletin* 56 (1959) 81-105.
 - [32] K. Schermelleh-Engel, Helfried Moosbrugger, & Hans Mueler Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures, *Methods of Psychological Research* 8 (2003) 23-74.
 - [33] J. Hair, .S., R.E. Tatham, R.L. , Black W.C., *Multivariate data analysis* Prentice Hall International, London, 1998.

ADAPTATION AND CONSTRUCT VALIDATION OF THE INDONESIAN VERSION OF THE UTRECHT WORK ENGAGEMENT SCALE

Yulia

Faculty of Psychology State Islamic University Syarif Hidayatullah Jakarta
yoel_anshor@yahoo.co.id

Abstract – This study uses confirmatory factor analysis to test a construct validity on adaptation scale of Utrecht Work Engagement Scale (UWES). The scale consists of three dimensions; vigor, dedication, and absorption. A total of 17 items which are vigor (6 items), dedication (5 items), and absorption (6 items). The subject were 52 employees of several working groups (teachers, collared employees, private employees, using convenience sampling technique). We used a procedure of adaptation as a cross-cultural adaptation by Beaton 2000. The results was six items are dropped because they are invalid; 1, 6, 7, 12, 15, 16. Item 1 and 6 are part of vigor dimension, item 7 is part of dedication, item 12, 15, and 16 are part of absorption dimensions. The results can be used as a scale for conducting research in the future. Then, the measuring instrument Utrecht Work Engagement Scale (UWES) can be used in business circles to determine the level of involvement of workers in the workplace.

Keywords: *Confirmatory Factor Analysis, Construct validity, Utrecht Work Engagement Scale.*

I. INTRODUCTION

Study in work engagement as a positive condition (Schaufeli et al, 2006). It is a quite and a lot of attention has been given to the topic - a topic such as effect of the recovery and proactive behavior is positively associated with work engagement (Sonnentag, 2003); personal resources and work is positively associated with work engagement (Bakker et al, 2007); job demand may strengthen positive associated resources with engagement. However, when job Demand is higher, then a positive relationship between Resources and engagement is weakened (Bakker, 2011); work life imbalance spoil a positive relationship between self-efficacy and work engagement (Virga et al, 2015), matter it is refers to a set of behavioral characteristics associated with the job. Author focused on how the Utrecht Work Engagement Scale (UWES) consists of 17 items with Cronbach higher than 0.80 with the analysis of three factors, namely vigor (6 items) refers to the work of passion and perseverance in the face of adversity, dedicated (5 items) refers to enthusiasm, pride, and challenge, absorption (6 items) refers to difficulty in separating themselves from the job (Schaufeli et al, 2004). Furthermore, this measure has a Likert scale of 7 points in the range of 0 - indicating "never", 1 - shows "never" with intensity "almost", 2 - indicating "rarely", 3 - show the "sometimes - sometimes" 4 - shows "often", 5 - showed "very often", and 6 - showed "always" then confirm factor analysis approach is used to show the structure of the three scale factors are superior in terms of fit to arrange engagement factor (Schaufeli et al, 2002). Zecca et al. (2014) to get the internal validity of the French people with an alpha value of 0.93, the subject of Brazil with an alpha value of 0.95 (Vazquez et al, 2015).

The purpose of this study was to assess the validity of Utrecht Work Engagement Scale (UWES) version of the 17 items in the Indonesian language by using the approach of confirmatory factor analysis (CFA) allows each item - the item is unidimensional (meaning that only measures one factor alone), so the Utrecht Work Engagement Scale of adaptation version Indonesian academics used for future research and business to determine the level of engagement of workers in the workplace.

II. METHOD

A. Participants

In collecting the data, initially researchers was distributed questionnaires to 100 employees of the factory of PT X. Because researchers pain Magh, researchers did not accompany and appealed to the chairman of the organization X and an employee of PT X to distribute questionnaires however, the questionnaire the researchers did not return and the researcher to drop out. Ultimately researchers in the aid of a friend with accidental sampling techniques, to anyone who is still actively working (employees), at around the researchers house with the address Highway Kalimulya Kebon duren RT.03 RW.05 Village Kalimulya Cilodong District of Depok City 16 413, including alfamart employees, a workshop, a bank clerk, PLN, tax consulting, English language course, and educators willing to fill out a questionnaire, in to get 52 participants.

B. Adaptation Procedure

Step - adaptation measures undertaken by the author by Beaton et al (2000), as follows:

1. Author are looking for a scale used by many countries and elected measuring devices work engagement that Utrecht Work Engagement Scale (UWES) developed by Wilmar Schaufeli.
2. Author asked for permission from Wilmar Schaufeli via email to use the gauge Utrecht Work Engagement Scale (UWES) as a college assignment to study the non - commercial of the measuring adaptation instrument.
3. Author do translation and back translation item that was helped by Teguh Fachmi a teacher of English and have a background in English language was good. This is due to retain its original meaning when translated into Indonesian version. Like translators judged that the original version of the item vigor (at my job, I am very resilient, mentally) demonstrate an understanding oblique term 'very resilient', because the items are tailored to 'at my job, I am very tenacious (diligent) ', then a customized version back, the Indonesian version into English as compared to the original UWES, no change much.
4. Author conducted item a discussion with two students majoring in psychometrics and have a good background in psychology.
5. Author conducted a discussions with some of the workers, namely, three teachers and three workers to understand the written statement of the items in the questionnaire.

C. Validity Construct

In order to test the construct validity of the instrument The Social Provisions Scale, author used a method CFA (confirmatory factor analysis). The CFA was conducted using software LISREL 8.70 and The logic of the CFA by Umar (2012):

1. There is a concept or trait that defined operationally so it can be compiled questions or statements to measure it. This trait is called a factor, while the measurement of these factors is done through an analysis of the responses to items.
2. Theorized each item only measures one factor as well. This means that either the item or subscale is unidimensional.
3. With the available data can be used to estimate the correlation matrix between the items that should be obtained if indeed unidimensional. The correlation matrix is called sigma (Σ), then compared with a matrix of empirical data, which is called the matrix S. If the theory is correct (unidimensional) then of course there is no difference between the matrix S - matrix Σ or can be expressed by $S - \Sigma = 0$,
4. The statement made null hypothesis is then tested by chi square. If the results are not significant $P\text{-value} > 0.05$, then the null hypothesis "not rejected". That is the theory unidimensional unacceptable that item only measures one factor alone.
5. If the model fit, then the next step to test whether or not a significant item measure what is to be measured, using a t-test. If the results of t-test was not significant (<1.96) then the item is not significant in measuring what is to be measured, if necessary, such items being dropped.
6. Furthermore, if the results of the CFA are items load factor coefficient is negative, then the item must be dropped. The item means to measure the opposite of what is to be measured. Nevertheless, it should be checked whether the item is an item in the form of negative (unfavorable). For items that are unfavorable prior CFA analysis performed.

III. RESULT

The Utrecht Work Engagement Scale (UWES) consists of 17 items, classified into three dimensions: the vigor (6 items), dedicated (5 items), and absorption (6 items) (Schaufeli et al, 2002). Based on the results of the CFA for each dimension, are:

Table 1. Summary of Items Test and Reliability

Component	Item No	Standard error	Coeffisien	T-test	Significant
Vigor	1	(0.15)	0.07	0.47	X
	2	(0.12)	0.87	7.32	V
	3	(0.12)	0.93	7.96	V
	4	(0.13)	0.63	4.83	V
	5	(0.14)	0.39	2.80	V
	6	(0.14)	0.27	1.90	X
Dedicated	7	(0.14)	0.10	0.70	X
	8	(0.12)	0.83	7.19	V
	9	(0.13)	0.72	5.67	V
	10	(0.11)	0.91	8.26	V
	11	(0.11)	0.91	8.10	V
Absorption	12	(0.13)	0.15	1.11	X
	13	(0.14)	0.42	2.95	V
	14	(0.15)	0.60	4.00	V
	15	(0.13)	0.23	1.70	X
	16	(0.13)	0.08	0.59	X
	17	(0.17)	1.07	6.16	V

From Table 1 above that the model of the factors theorized by the Utrecht Work Engagement Scale (UWES) on items 2, 3, 4, 5, 8, 9, 10, 11, 13, 14, 17 shows the results of t-test ($>1,96$) then the item is significant in measuring what is to be measured, such items are accepted and items 1, 6, 7, 12, 15, 16 shows the results of t-test ($<1,96$) then the item is not significant in any measure to be measured, if necessary, such items are rejected.

Table 2. Model Fit

Component	Chi - square	df	RMSEA (90%)	P
Vigor	13.31	9	0.000	0.148
Dedicated	8.59	4	0.000	0.072
Absorption	10.30	8	0.000	0.244

From Table 2 above researchers tested whether each component factors unidimensional, are means it only measures one factor. Then, the acquired model fit and obtained the value of p-value > 0.05 states that a model with one factor may be acceptable.

IV. CONCLUSION

The result is 11 of 17 items is valid. The results was six items are dropped because they are invalid; 1, 6, 7, 12, 15, 16. Item 1 and 6 are part of vigor dimension, item 7 is part of dedication, item 12, 15, and 16 are part of absorption dimensions. Based on the results, that work engagement is universal in its definition even though there are items due to cultural differences deserve to be taken into account.

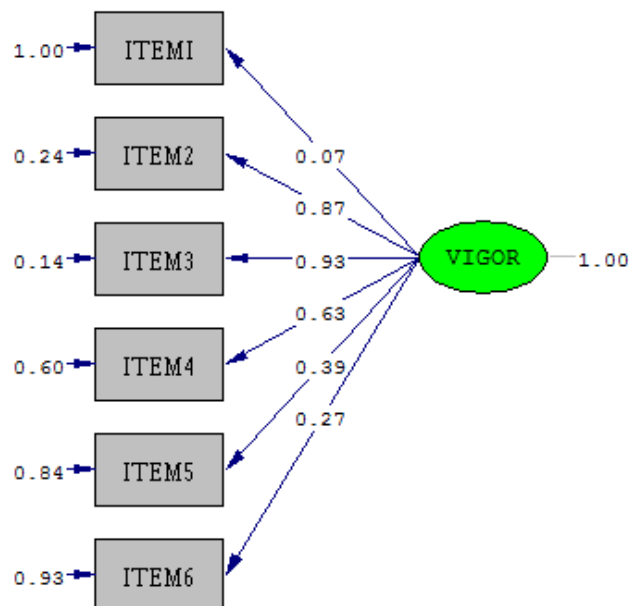
The constraint of this study is small sample and have not covered various types of work. For the future study, sex and age could be included to elaborate the study.

REFERENCES

- [1] Bakker, B, Arnold,. (2011). An Evidence-Based Model of Work Engagement. Journal Permissions. 20:265

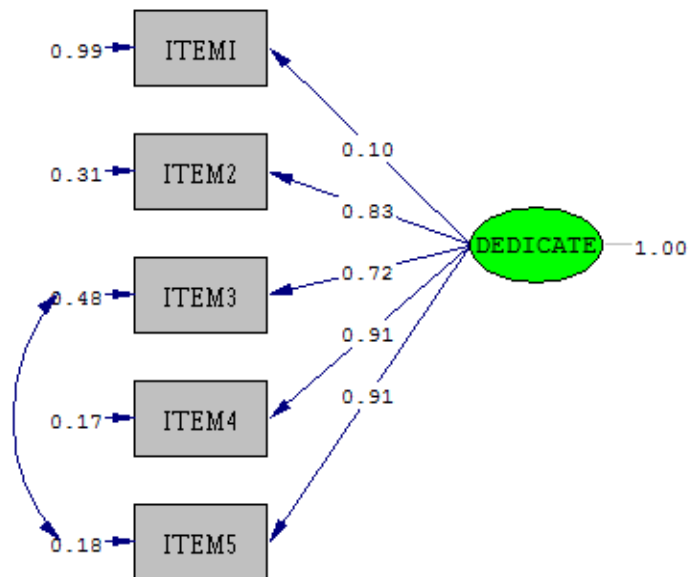
- [2] Bakker, B, Arnold., Demerouti, E., Hakanen, J, J., Xanthopoulou, D,. (2007). Job Resources Boost Work Engagement, Particularly When Job Demands Are High. *Journal of Educational Psychology*. Vol. 99, No. 2, 274-284.
- [3] Beaton, D, E., Bombardier, C., Guillemin, F., & Ferraz, M, B. (2000). Guidelines for the process of cross-cultural adaptation of Self-report measures. *Spine*, 25(24), 3186-91.
- [4] Schaufeli, W. B., Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*. 25. 293–315.
- [5] Schaufeli, W. B., Bakker, A. B. Salanova, M. (2006). The Measurement of Work Engagement With a Short Questionnaire A Cross-National Study. *Educational and Psychological Measurement*. Vol. 66, No. 4, 701-716.
- [6] Schaufeli, W., Salanova, M., Gonzales-Roma, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two simple confirmatory factor analytic approach. *Journal of Happiness Studies*, 30, 71–92.
- [7] Sonnentag, S,. (2003). Recovery, Work Engagement, and Proactive Behavior: A New Look at the Interface Between Non-work and Work. *Journal of Applied Psychology*, 88, 3, pp. 518-528.
- [9] Umar, J. (2012). Confirmatory factor analysis: Bahan ajar perkuliahan. Jakarta: UIN Syarif Hidayatullah.
- [10] Vazquez, S. C. A., Magnan, S. E., Pacico, C. J., Hutz, S. C., & Schaufeli, W. (2015). Adaptation and Validation of the Brazilian Version of Work Engagement Scale. *Psico-USF, Braganca Paulista*, v. 20, n. 2, p. 207 – 217.
- [12] Virga, D., Horga, A., Illiescu, D,. (2015). Work–Life Imbalance as a Moderator in the Relationship Between Resources and Work Engagement. *Journal of Personnel Psychology*, Vol. 14(2):80–90.
- [13] Zecca, G., Gyorkos, C., Becker, J., Massoudi, K., Bruin de, P, G., Rossier, J. (2015). Validation of the French Utrecht Work Engagement Scale and its relationship with personality traits and impulsivity. 19-28.

V. APPENDIX



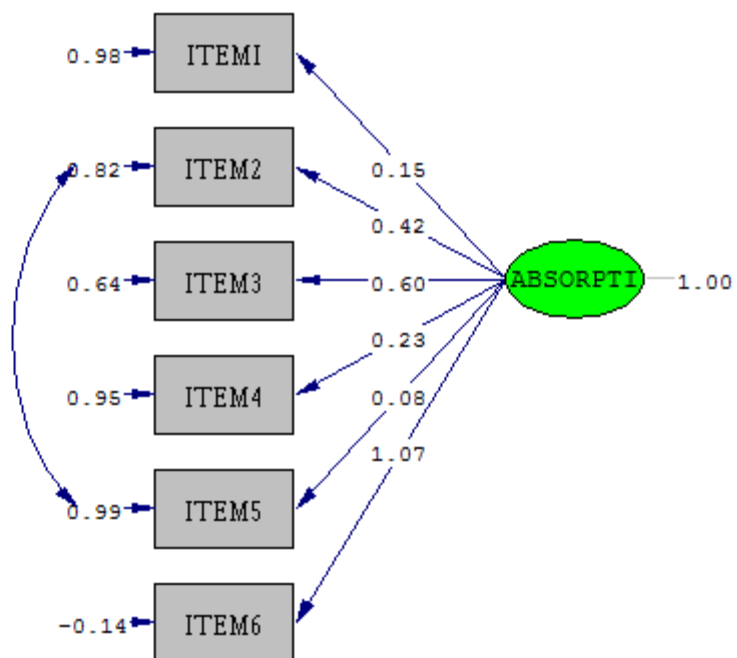
Chi-Square=13.31, df=9, P-value=0.14893, RMSEA=0.000

Figure 1. Path Diagram Vigor



Chi-Square=8.59, df=4, P-value=0.07223, RMSEA=0.000

Figure 2. Path Diagram Dedicated



Chi-Square=10.30, df=8, P-value=0.24449, RMSEA=0.000

Figure 3. Path Diagram Absorption

THE EFFECTIVENESS OF THE BOARDING TEACHER PROFESSIONAL DEVELOPMENT PROGRAM: AN APPROACH OF PROCESS EVALUATION

Friyatmi

Doctoral Student at PPs UNY & Lecturer at UNP
fri.yatmi@gmail.com

Abstract—This study aimed to evaluate the effectiveness of the implementation of the boarding Teacher Professional Development Program (PPG SM-3T) in State University of Padang (SUP). This research was an evaluation study using a part of the CIPP model, namely the process evaluation. The research questions for this study were 1) how is the effectiveness of the implementation of PPG SM-3T boarding program? 2) what are the weaknesses in implementing the PPG SM-3T boarding program in SUP? Data in this study were collecting using questionnaires and interview techniques. Informants of this study were the participants of the PPG SM-3T PSU and managers program. The data were analyzed using descriptive statistic techniques. The results of this study revealed that all over management of program implementation was considered less effective by the participants. The weaknesses of the program implementation are as follows. 1) The boarding education programs was less-organized, causing some programs were not function properly and often the schedule were not followed strictly; (2) Lack of coordination between the management and weakness supervision/controlling manager on the boarding activities resulted in less well executed program. (3) Meals service agent were less professional, resulting in the low quality and less variation food being served to the participants.

Keywords: *Teacher profession, PPG, boarding school for teachers*

I. INTRODUCTION

Teacher quality plays an important role in educational process. Several efforts have been made by the Indonesian government to improve the teachers quality. Reference [1] reveals that the effort might be giving training, rotating, transferring, performance management and teachers career development programs. Although the government has made various effort, some research on teachers performance indicate that teachers knowledge, skills and performance are below the teacher competency standards in many areas of Indonesia [2] [3]. Moreover, some research showed that the teacher's certification program has not positive impact on improving teachers performance [4, 5].

To reduce the problem, the new policy has established by the government by issuing the government regulations about teacher professional development program. This program call Pendidikan Profesi Guru (PPG) which is a pre-service teacher education program for teachers candidat. After participating in the PPG programs, it is expected that the teacher candidates develop various professional teaching competencies. In addition, teachers are also develop soft skills such as environmental awareness, mental discipline, cooperative, and honest attitudes. International experience has shown that the implementation of teacher professional development program as a postgraduate program will provide the professional teacher [4].

The PPG programs are held by Lembaga Pendidikan Tenaga Keguruan (LPTK) which are teacher training institute in several universities appointed by Indonesian government. There are several type of PPG programs that have been implemented by the government, such as Integrated PPG, Collaborative PPG, and PPG SM-3T. Among those PPG programs, the PPG SM-3T is attended by many teachers candidates. SM-3T is named for a program of *Sarjana Mendidik di daerah Terluar, Terdepan dan Tertinggal*. It is a year teaching program for graduates in the Frontier, Outermost, and Disadvantaged Regions of Indonesia. To be involve in PPG SM-3T, a teacher candidate is required to participate in SM-3T program for one year.

The purpose of PPG SM-3T program is to produce professional teacher candidates who have four competencies, i.e pedagogy, professional, personal, and social competence. The first two competencies are trained in workshops activities on campus, while the last two competencies fostered through boarding program.

As a boarding program, all participants of the PPG SM-3T program are stay and interact each other in a dormitory house for one year. The main goal of this program is to develop the personal and social competence of the participants. To achieve these objectives, facilities, instructors, curriculum, and learning materials are prepared thoroughly. The participants are requested to actively involved in various activities designed and facilitated in the boarding house. Various boarding programs have been designed to enhance the personal and social competence of the teacher candidats, including religious activities, tracing the individual talent, attitudes, aptitude and interest of participant, sports, arts performance, social and cultural activities, and mental development. The PPG SM-3T boarding program is more eager to highlight the mastery of personal and social competence so that might be expected to produce a professional and qualified teachers. The PPG SM-3T boarding program has been implemented since 2013 and has succeeding in its outcome in professional teacher candidates who have been certified as profesional educators.

This program is funded by the national government, with an expectation that in the long time, the program would produce prospective teachers with noble character and professional competence. A boarding training and education system requires a substantial amount of funding from the government.

Boarding shool programs is not a new concept in the context of Indonesian education system. Boarding education has been widely implemented in primary and secondary schools, ranging from Islamic traditional boarding school well known as pondok pesantren, to modern boarding secondary senior school as well as other religious based schools system. This boarding school concept was adopted to foster the professional teacher candidates in the form of PPG SM-3T boarding program. A concern is being made, as the participants of this boarding education programs are adult teachers candidate, who are considered to be able to develop their own character. In fact, those participant of SM-3T program already had their experience of teaching in challenging areas and far away from their families, but they were able to survive. They were able to socialize with the community and think creatively in facing various limitation in the frontier area.

An evaluation is needed to analyzed the effectiveness of the boarding PPG SM-3T program. It is necessary to find the weakness of various aspects of the program, then it can be easily to perform improvement treatment. This research aimed to evaluate the process of PPG SM-3T boarding program at State University of Padang (SUP). The problem raised are (1) how is the effectiveness of the implementation of PPG SM-3T boarding program organized by SUP? (2) what are the weaknesses in implementing the PPG SM-3T boarding program organized by SUP?

II. RESEARCH METHODS

This research utilized part of the CIPP evaluation model (Context, Input, Process and Product), where the emphasize is in the process evaluation of the PPG SM-3T boarding program. Sample for this research was selected by proportional random sampling technique from paticipants and the managers of boarding program who participate in the 3rd batch of PPG SM-3T year 2015 at SUP. Data was collected through questionnaire and interview. The questionnaire was administered to evaluate the process of PPG SM-3T boarding program. Interview was conducted to the participants and program managers to analyze the constrains or obstacles in implementing the process of boarding program at Padang State University, West of Sumatra. The effectiveness of the implementation of PPG SM-3T boarding program was assessed through following indicators: 1) the managing of program implementation, 2) the strategy of program implementation, 3) developing of personal competence program, and 4) developing of social competence program. The data was analyzed using statistic descriptive analysis. The evaluation findings of this research was compared to the prescribed evaluation criteria. The criteria are follow, 1) score $X \geq 3$ shows a high criteria (effective), 2) score $2 \leq X < 3$ means moderate categori (less effective), and 3) score $X < 2$ shows poor criteria (not effective).

III. RESEARCH FINDING AND DISCUSSION

A. The Effectiveness of The Process of PPG SM-3T Boarding Program

The finding of this study indicate that the process of PPG SM-3T boarding program for the year 2015 at SUP were rated in the moderate category as shown in Table 1 below.

Table 1. Results of Evaluation Process of PPG SM-3T Boarding Program

No	Component	Average	Category
1	Managing of program implementation	2.31	Moderate
2	Strategy of program implementation	2.37	Moderate
3	Developing of personal competence program	2.49	Moderate
4	Developing of social competence program	2.86	Moderate
		2.51	Moderate

The PPG SM-3T boarding program is intended to prepare teachers candidate to become professional teacher and having personal dan sosial competence. In this study, although the implementation of both programs are in the moderate category, but the score social competence development higher than other indicators. The interaction among the participants was in good category, with average score of 3.14, while the interaction between the participants with the boarding manager is at moderate category as like as the implementation of the development program for personal skills i.e the conveying of the discipline, attitude, creativity, responsibility, spiritual & character development programs.

Character building is an important key in the implementation of PPG SM-3T boarding programs. Therefore, practice and habituation are two main principles introduced in that programs. The participants were trained to be obedient to the guiding norms and habituating those norms in their daily life. These principles generally are applied in all group activities in their boarding life. These practice and habituation are expected to become culture in the daily life. Reference [6] opined that one strategy leading into character development is through habituation. Habituation may lead behaviour become permanent character. Based on the observation to the participants of the PPG SM-3T, these principle had been implemented and habituated among them. For example, close and intense interaction among the participants could be seen in all group activities, e.g praying, eating, learning together, despite they come from diverse region, culture, and custom from all region in Indonesia. The concept of boarding school has a significant influence on the social competency development of the students because the boarding life can be used as a tool to improve the spirit of brotherhood and cultural identity [7]. Participants of PPG SM-3T were coming from a very diverse social and cultural backgrounds, the boarding life develop close relationship among the participants and also they learn about others culture. These skill are essential for the teachers to prepare them entering and adapting to their working place. It is very important for the PPG participants to support their duties and responsibility as teachers, after they graduated from the program, and return to their respective rural areas in Indonesia.

Implementation of various boarding program can improve the spirituality, mental health, talents and interest of participants. Based on the researcher observation, the Moslem participants held religious studies and discussion to increase their spirituality regularly. The similar program also given to other participants who belongs to other religious denomination. In addition, art and cultural performance also given every Saturday to accommodate and develop participants respective talents and interest.

The lack of the effectiveness of the program implementation caused by the the strategy of the program has not been maximized. The method of program implementation and its compliance with the program schedule was also rated moderate by the participants with average scores respectively are 2.42 and 2.40. The lowest score of the implementation of boarding was at the performance of instructors/speakers with the average of 2.36.

Although the efectiveness of the implementation of boarding program was rated moderate, the managing of program implementation were rated lower than other indicators. That related to consumption management, environmental hygiene & security in dorm, and the supervision of the

manager. based on that four aspect, consumption management is at low category while the others were at moderate category.

The objective of boarding program is to buliding character among the teachers candidate with high achievement, mastering life skills, and enjoying healthy life. To achieve that objective, participants are provided with dormitory equipped, facilities, and basic needs, including healthy food. Participants are mandatory to get three times daily meals. Research finding indicated that the quality of meals being served daily at the dormitory was at low category with score 1.9. Based on the interview with boarding management, the catering agency selected were less professional. Less food variations, less food quality and less tasty meals served daily lead into dissatisfaction of the participant. Research [8] revealed that boarding education has benefits for students to have a positive life patterns, which is supported by a nutritious food. The committee should improve the consumption management to ensure the nutrition and the healthy of participants.

Participant also considered the lack of supervision by the program administrator and boarding house management as well as the lack of evaluations done by the administrator. Interview with the boarding management indicated that the evaluation on the boarding life was not done regularly, and it involved only limited participants who were selected as boarding committee members. It intended to increase the effectiveness of evaluation process. The existence of evaluation is very important in an program because it can beneficial to the development and quality program improvement [9].

B. The Weaknesses in Implementing The PPG SM-3T Boarding Program

In addition to administered questionnaire, a number of closed questions were given to participants and boarding managers of PPG SM-3T to identify the weaknesses of the program. Based on the answer to those questions by the participants and managers, in general the shortcomings of the PPG SM-3T boarding program were focused on the following points: 1) The boarding programs was less-effective, causing some programs were not function properly and often the schedule were not followed strictly; 2) Lack of coordination between the management and weakness supervision/controlling manager on the boarding activities resulted in less well executed program. 3) Meals service agent were less professional.

The finding from evaluation process of PPG SM-3T boarding program organized by Padang State University indicated that it was not up to the maximum success, due to several factors. These factors were not interrelated each other, neither they would affect the other components of the boarding program. Based on the research finding, these factors were lacking of coordination among the members of the boarding program organizers, lack of supervision, and unsatisfactory food services.

The main foundation for a successful boarding program is the availability of team works who are able to design and supervise the whole program according to the desired goal. In this respect, program planning play the most important role. Communication and coordination amongst the program organizer need to be maintained strictly, to ensure the whole component of the program functioning properly.

IV. CONCLUSION

The goal of PPG SM-3T is to increase the teachers quality by developing four teacher competencies which are (1) pedagogical competency; (2) professional competency; (3) social competency; and (4) personality competency. The emphasize of the boarding program is the development of social competency and personality competency by close interactions and association among the participants while they stay in the dormitory for one year.

Evaluation on the process of PPG SM-3T of SUP boarding program indicated that all over management of program implementation was considered less effective by the participants. The weakness of the program include the following areas: (1) The boarding education programs was less-organized, causing some programs were not function properly and often the schedule were not followed strictly; (2) Lack of coordination between the management and weakness supervision/controlling manager on the boarding activities resulted in less well executed program. (4) Meals service agent were less professional, resulting in the low quality and less variated food being served to the participants.

Based on the research finding, recommendation are given to improve the quality of the PPG SM-3T boarding program in the future. The program organizer should improve their managerial

system by implementing the planning, organizing, actuating, and controlling functions. More specifics recommendation are given for the following areas: (1) There is a need to improve the role of boarding management committee members by establishing job description and responsibility of each program components, based on the guidelines given by the Directorate of Higher Education, Minister of Education and Culture of Indonesia; (2) To recruit more professional and more than one catering agents to enable more varieties of food being served to the participants; (3) To implement a continuing supervision system to control the discipline of participants while attending the PPG SM-3T boarding program.

REFERENCE

- [1] Iswari, M., *Membina Perkembangan Emosi Remaja Meningkatkan Profesional Guru*. Pedagogi, Jurnal Ilmiah Ilmu Pendidikan, 2009. **IX**(1).
- [2] Tanang, H. and B. Abu, *Teacher Professionalism and Professional Development Practices in South Sulawesi, Indonesia*. Journal of Curriculum and Teaching, 2014. **3**(2): p. 25-42.
- [3] Bank, T.W., *Transforming Indonesia's Teaching Force*. 2010, The Word Bank
- [4] Chang, M.C., et al., *Teacher Reform in Indonesia: The Role of Politics and Evidence in Policy Making*. 2014, The Word Bank: Washington, DC
- [5] Hurmaini, M., *Dampak Pelaksanaan Sertifikasi Guru terhadap Peningkatan Kinerja Guru dalam Proses Pembelajaran: Studi pada Madrasah Tsanawiyah Negeri Kota Jambi*. Media Akademika 2011. **26**(4).
- [6] Miftah, M., *Pengembangan Karakter Anak Melalui Pembelajaran Ilmu Sosial*. Jurnal Pendidikan Karakter, 2013. **III**(2).
- [7] Graham, A.G., *The Power of Boarding Schools, A Historiographical Review*. American Educational History Journal, 2012. **39**(2).
- [8] Curto, V.E. and R.G. Fryer, *The Potential of Urban Boarding Schools for the Poor: Evidence from SEED*. Journal of Labor Economics, 2014. **32**(1): p. 28.
- [9] Stavropoulou, A. and T. Stroubouki, *Evaluation of Educational Programmes – The Contribution of History to Modern Evaluation Thinking*. Health Science Journal, 2014. **8**(2).

THE EFFECT OF FORMATIVE TEST TYPES AND ATTITUDES TOWARD MATHEMATICS ON LEARNING OUTCOMES

Hari Setiadi, Sugiarto, Rini

Universitas Muhammadiyah Prof.D.R. HAMKA (UHAMKA)

Abstract—This study aimed to analyze the effect of formative test types and attitude towards Mathematics on learning outcomes. The hypotheses are: (1) The outcome of students' Mathematics learning measured by multiple-choice test items is higher than that on essay test items. (2) There is an interaction effect between the formative test types and attitudes towards mathematics on learning outcomes. (3) The learning outcome of students who have a positive attitude on multiple choice test items is higher compared to those receiving essay test items. (4) The learning outcome of students who have a negative attitude on multiple choice test items is lower compared to those receiving essay test items.

Quasi-experimental method was used in this study. Data were analyzed using descriptive and inferential statistics. The sample in this study was 180 Grade 10 students in a government secondary school named SMAN 65, selected by cluster random sampling method. The results showed that: (1) The mathematics learning outcome measured by a multiple choice test was generally higher than that of an essay test, (2) There was an interaction effect between the formative tests and attitude towards Mathematics and the learning outcome in Mathematics, (3) The outcome of Mathematics learning of students who had a positive attitude towards Mathematics was higher compared to those receiving essay test questions. (4) The outcome of Mathematics learning of students who had a negative attitude towards Mathematics was lower compared to those receiving essay test questions. The implications of the study is discussed.

Keywords: *formative test, multiple choice test, essay test, experimental methods, Math learning outcome*

I. INTRODUCTION

Learning and teaching is a process of interaction between teachers and students. Successful learning activities is not only determined by the ability of teachers in applying teaching methods, techniques, media and tools used, but also by the evaluation of teaching and learning processes. By conducting appropriate evaluation methods, the teacher will be able to monitor the development of learning outcomes. A learning process can be said to be successful if the delivery of the subject matter of the teacher can be understood and mastered by students. The students' level of mastery of the learning materials can be seen from the results of formative tests conducted. If results of formative tests low, it means learning that takes place can be said to be less successful, and vice versa.

Multiple choice test has some advantages. First, the stem is short and students do not need to write the answer. Second, many items representing the topics learned could be included in test. Third, as an objective test, multiple choice test is considered more reliable than essay test that is subjective in nature. Fourth, multiple choice test can also be administered to a large number of students and the scoring can be done easily and quickly.

Subjective tests, often known essay, is one form of test questions in which scoring is influenced by opinions/ ratings. Essay test requires students to formulate their own answers. So students cannot choose an answer, they must answer in their own words. Essay scoring must also be done by an expert in the field of study. Essay test is used to fully develop the students' ability to respond to the questions given. In addition to memory and application of a concept, analysis and interpretation is necessary to answer the essay test. With essay test, teachers can determine whether a student has mastered a concept and the extent to which he has performed

an analytical ability. It is apparent from the students' answers written in the answer sheet. Each step in answering the question can be used as an indicator of the extent of student mastery.

According to Stenberg and Grigorenko (2009), multiple-choice test and essay items require three kinds of thinking ability, namely analytical, creative and practical in three domains of problems: verbal, math, and figural. Therefore, the low student learning outcomes, as the evidence of competence, can be attributed to the teacher's inability to create joyful, yet effective, learning activities.

There are some aspects to consider as the foundation of this research, including the difficulty and fear of students' learning of mathematics. Attitudes of students towards learning of mathematics may also affected the achievement of learning outcomes. Russefendi (1989) states that Mathematics courses are generally disliked and even hated by students. The negative attitude of such students leads to the low mastery of basic competencies in mathematics and in turn have an impact on the achievement of learning outcomes. A positive attitude should continuously be developed through life skills, where practice activities will reinforce emotional intelligence, identity, sense of belonging and independence. The attitude of the students tends to be a key success factor for the success in learning Mathematics. Thus, the need for research to investigate the relationship between the administration of formative test questions and students' attitudes toward learning mathematics, so that in the long run it might be expected to be able to improve students' mathematics learning outcomes.

A. Research Questions

1. Is there a difference between the Mathematics learning outcome of students who work on multiple choice questions and that on essay questions?
2. Is there an interaction effect between formative tests and attitude towards Mathematics?
3. Is the outcome of Mathematics learning of students who had a positive attitude towards Mathematics higher than those working on essay test questions?
4. Is the outcome of Mathematics learning of students who had a negative attitude towards Mathematics lower than those working on essay test questions?

II. RESEARCH METHODOLOGY

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size, having 4-3-4-3 margin. The paper is written in English not exceeding 12 pages.

A. Methods

The method used in this study was an experimental research. According to Sudjana (1985), the goal of the experimental research is to compare and investigate whether there is a causal relationship as well as the extent of the causal relationship by providing treatment in the group being investigated. However, this study used a quasi-experiment research, which is not pure experiment (Sukmadinata, 2005). This is due to the impossibility of variables that can be controlled and affect the results of the study. Experimental study in this research was conducted by administering multiple choice test to the experimental group and essay test to the control group.

B. Research Design

This study used Treatment by Levels Design. According to Hadi (2004) such a method is used due to the stratification and levels in study subjects. As non-experimental variables or non-treatment variables, the stratification must be controlled by equalizing them. The study design is described as follows:

Table 1. Experimental Research Design

Attitude (B)	Mathematics Formative Assessment		
	Multiple Choice Test	Essay Test	$\sum b$
Positive	A_1B_1	A_2B_1	B_1
Negative	A_1B_2	A_2B_2	B_2
$\sum a$	A_1	A_2	AB

Notes:

- A_1 : Students' average score on multiple choice test
- A_2 : Students' average score on essay test
- B_1 : Students' average score on positive attitude toward Mathematics
- B_2 : Students' average score on negative attitude towards Mathematics
- A_1B_1 : Average score of students having positive attitude on multiple choice test
- A_2B_1 : Average score of students having positive attitude on essay test
- A_1B_2 : Average score of students having negative attitude on multiple choice test
- A_2B_2 : Average score of students having negative attitude on essay test

1. Population

The population is divided into two categories: target population and accessible population. The target population is the population that reasonably has similar characteristics with accessible population. Meanwhile, accessible population or population that is measurable in reality and used as the basis from which the sample is drawn. The target population in this study was all 569 students at SMAN 65 enrolled in the academic year 2010-2011, whereas the experimentally accessible was all students in Grade X SMAN 65 totaling 180 students.

2. Sample

In order to obtain a representative sample, this study used the cluster random sampling technique, that is, the selection or determination of the sample by groups. The following is the spread of the sample in this study.

Table 2. Subject in Treatment Group

Attitude	Mathematics Formative Assessment		
	Multiple Choice Test	Essay Test	$\sum n$
Positive	13	13	26
Negative	13	13	26
$\sum n$	26	26	52

Table 2 above shows that the students who were given multiple choice test and those who were given essay test being the treatment groups. Each group consists of 26 students; each group is evenly divided into two subgroups, namely those with positive attitude and negative attitude.

3. Data Analysis Technique

There were three steps conducted in analyzing the data of the study, namely (1) describing the data of the research variable, (2) examining the requirements for analysis, and (3) testing the hypothesis.

The first step was conducted using Two-way ANOVA followed by the Tuckey's t-test to assess whether the means of the two groups are statistically different from each other. The second step involved tests of normality and homogeneity assumptions. This step is necessary in comparing two groups using Two-way ANOVA (Tabachnick & Fidell, 2001). The test of such assumptions can use the Kolmogorov-Smirnov test

III. RESEARCH RESULT

The research result is shown in Table 3 below.

Table 3. Summary of Research Result

Formative Test Types Attitude	Formative Test (A)		Total
	Multiple Choice Test (A ₁)	Essay (A ₂)	
Positive (B ₁)	n = 13 \bar{x} = 88,69 s = 5,089	n = 13 \bar{x} = 75,92 s = 5,299	n = 26 \bar{x} = 82,31 s = 8,264
Negative (B ₂)	n = 13 \bar{x} = 65,31 s = 5,360	n = 13 \bar{x} = 71,38 s = 7,171	n = 26 \bar{x} = 68,35 s = 6,934
Total	n = 26 \bar{x} = 77,00 s = 12,977	n = 26 \bar{x} = 73,65 s = 6,520	n = 52 \bar{x} = 75,33 s = 10,331

A. Hypotheses Testing

1. Learning outcome of students who received multiple choice test and that of students who received essay test

The students' average score on multiple choice test was 77.00, while the other on essay test was 73.65. The scores show the difference in Mathematics learning outcomes of both groups, where the score of students working on multiple test was higher than that of students working on essay test. Based on an F-test, the difference was significant.

2. The interaction effect between formative test types and attitudes toward Mathematics

The results show that there was a significant interaction effect between the formative test types and the attitude toward Mathematics as depicted in Figure 1 below.

Estimated Marginal Means of Hasil Belajar Matematika

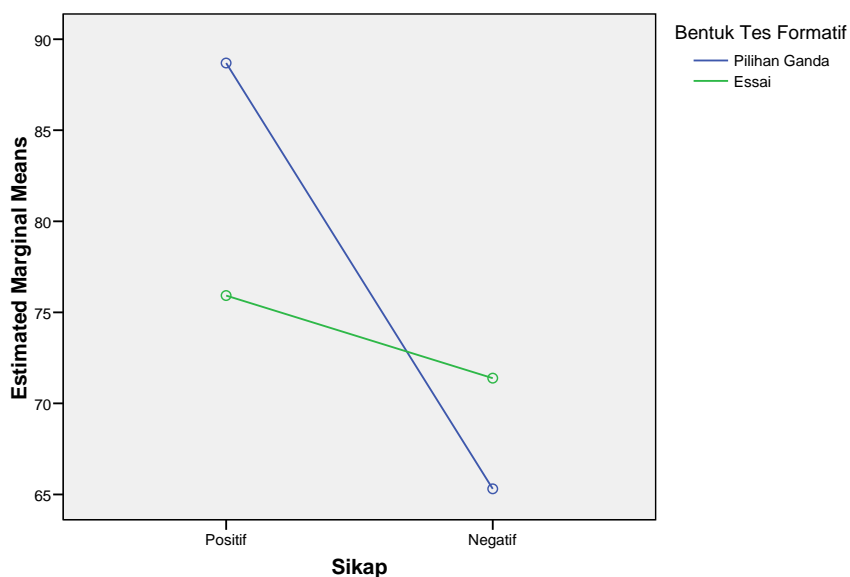


Figure 1. The Interaction of the formative test types and the attitude toward Mathematics

3. The outcome of Math learning of students who have positive attitude and given multiple choice test is higher than that of students given essay test
4. The outcome of Math learning of students who have negative attitude and given multiple choice test is lower than that of students given essay test

IV. DISCUSSION

A. Learning outcome of students who did Math multiple choice items was higher than those who did essay items

The data analysis suggests that the hypothesis is supported in that the mathematics learning outcomes of students who received multiple-choice test questions are higher compared with those receiving essay items. These results are in line with the statement by Silverius (1991) that essay tests require the student's ability to express, organize, and integrate the ideas using people's own words. Essay tests require students to think comprehensively. Students who lack mastery of the subject matter will have difficulties and likely to earn low learning outcomes. The results of this study are also consistent with the statement made by Tola (2007) that multiple-choice test questions require answers that can be selected from several possible answers. The availability of answers is helpful for students to find the right answer. This will make students have a greater chance to answer correctly than those who are given essay tests.

B. Effect of interaction between formative test forms and Attitudes toward Math learning outcomes

The analysis showed that there is a significant interaction effect between forms of formative test questions and attitude towards mathematics student learning outcomes. These results are in line with Tessmer's statement (1993) that formative tests are constructed to measure the extent to which a particular passage on one subject or sub-subject is already occupied by students. Formative test forms and attitudes of both variables equally affect the results of students' mathematics learning. The results of this study are also consistent with the statement by Morgan (in Sternberg, 2009) that attitude as their tendency to react either positively or negatively to a person, object or situation where Math is as an object of the attitude of the students. If mathematics is considered negative, regarded as a difficult subject and does not have the value of the benefit, it will make students less loved and tended to avoid. This can have an impact on students' mathematics learning achievement. In contrast, a positive attitude will make students perceive positively also on Mathematics. They will miss if they are not present in the classroom and lead the students to be responsible for learning. This positive attitude resulted in increased student learning outcomes in Mathematics. Thus, it is evident that there is an interaction effect between the formative test forms and the attitude towards Mathematics student learning outcomes.

C. Learning outcome of students who have a positive attitude and did multiple choice tests is higher than that who did essay test

The analysis showed that the hypothesis is verified that Mathematics learning outcomes of students who have a positive attitude and did multiple choice test items is higher than that received essay items. These results are in line with Surapranata (2004) that multiple choice tests have the advantages in measuring various aspects of cognitive levels, ranging from memory to the evaluation. In contrast, the narrative form has the disadvantage that it is impossible to test all the subject matters. The positive attitude encourages students in an attempt to achieve the mastery, knowledge, understanding, skills and abilities in learning Mathematics. Students will miss if he does not attend the lessons because they will lose the opportunities of mastering the competencies that should be obtained from the lesson. A positive attitude makes students responsible for learning by doing any assignment from teachers. Giving multiple-choice test questions regularly will make students skillful in working on the problems that can measure the mastery of all cognitive aspects of the subject matter, ranging from the level of memory or knowledge to the evaluation. In other words, It improves the students' overall cognitive domains.

Essay items will improve the students' analytic ability but they will have some difficulty in dealing with tests that measure broad range of learning materials. This will make learning outcomes lower. This evidence suggests that the students who have a positive attitude

Mathematics learning outcomes receiving multiple choice test items reach higher learning outcomes compared with that receiving essay test items.

D. Mathematics learning outcome of students who have a negative attitude towards multiple choice tests is lower than that of essay tests.

The analysis showed that the hypothesis is verified in which Mathematics learning outcome of students who have a negative attitude towards multiple choice tests is lower than that of essay tests. This result is consistent with Tola (2007) that students who have a negative attitude may have been misled by distractors since they do not have mastery and ability comprehensively. Therefore, when given a multiple-choice test questions they tend to guess because they do not have enough preparation. Guessing factors and reluctance to learn cause them reach lower learning results.

Thus, it is evident that the students who have a negative attitude towards multiple choice test items is lower than that of essay test items. Students tend to rely on guessing without having a good mastery of the materials. In contrast, students who were given essay test items tend to get highly motivated in learning and mastering the subject matter. This is because they are aware of their capabilities and limitations so that when they are facing formative tests they prepare the study and do not rely on guessing.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

1. Mathematics learning outcome of students who were given multiple choice test items is generally higher than that of essay test items;.
2. There is a significant interaction effect between multiple choice test items and essay test items with an attitude towards mathematics learning outcomes. Thus, applying test items appropriate to the learning attitude of the students will be able to improve learning achievement;
3. For students who have a positive attitude, Mathematics learning outcome on multiple choice test items is higher than that of essay test items;
4. For students who have a negative attitude, Mathematics learning outcome on multiple choice test items is lower than that of essay test items.

B. Suggestions

1. Teachers need a lot of practice making formative tests in the form of multiple choice of good quality. This is because multiple choice test can be used to cover a broad range of subject matter;.
2. Teachers should encourage students to develop positive attitudes towards Mathematics in particular, and other subject matters in general.

REFERENCES

- [1] Gunawan. No. 06/Th. XV/2010. *Membangun Sikap Positif terhadap Bahasa Indonesia*. Jakarta : Majalah Gema Widyakarya.
- [2] Abror, Abd. Rachman, 1993. *Psikologi Pendidikan*. Yogyakarta: Tiara Wacana Yogya.
- [3] Ahiri, Jafar. 2008. *Teknik Penilaian Kelas dalam Pembelajaran*. UHAMKA Press, Jakarta. Cetakan pertama.
- [4] Arikunto, Suharsimi. 2001. *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara, Jakarta. Asmin. 2006. *Pengaruh Ragam Bentuk Tes Objektif dan Gaya berpikir*. Jurnal Pendidikan dan Kebudayaan: No. 062, Tahun ke-12,.
- [5] Azwar, Syarifuddin. 1998. *Sikap Manusia: Teori dan Pengukurannya*. Yogyakarta: ANDI Offset.
- [6] Depdiknas. *Pengembangan Perangkat Penilaian Afektif*. Jakarta. 2010
- [7] Djaali & Pudji Muljono. 2008. *Pengukuran dalam Bidang Pendidikan*. Jakarta: Grasindo
- [8] Fathurohman, Pupuh & M. Sobry Sutikno. *Strategi Belajar Mengajar Melalui Konsep Umum dan Konsep Islami*. PT. Refika Aditama. Cetakan ke dua, 2007
- [9] Gerungan, W. A. 1996. *Psikologi Sosial*. Bandung: PT Erisco.
- [10] Grounlund, Norman E. 1981. *Menyusun Tes Hasil Belajar*, Terjemahan Bistok Sirait, Medan: Dirjen Dikti Departemen Pendidikan dan Kebudayaan.
- [11] Hall, C. S. and Lindzey, G. 1981. *Theories of Personality*. Singapore: Willey and Son Inc
- [12] <http://file.upi.edu/Direktori/C%20%20FPBS/JUR.%20PEND.%20BAHASA%20DAERAH/195901191986011%20%20USEP%20KUSWARI/Bab%20VI%20Tes%20Hasil%20Belajar%20Bahasa.pdf.2010>
- [13] <http://indramunawar.blogspot.com/2009/06/hasil-belajar-pengertian-dan-definisi.html>
- [14] <http://tarmizi.wordpress.com/2009/03/08/komponen-pembentukan-sikap-belajar-siswa>

- [15] Iskandar. Maret 2009. *Psikologi Pendidikan Sebuah Orientasi Baru*. Banten : Gaung Persada
- [16] *Kamus Besar Bahasa Indonesia*, 2007. Jakarta: Balai Pustaka.
- [17] Karmel, Louis J. *Measurement and Evaluation in the schools London: the Macmillan company*, 1970
- [18] Karso. 1993. *Dasar – dasar Pendidikan MIPA*. Jakarta: Universitas Terbuka.
- [19] Majid, Abdul. 2008. *Perencanaan Pembelajaran Mengembangkan Standar Kompetensi Guru*. Bandung: Remaja Rosdakarya cetakan kelima.
- [20] Mar'at. *Sikap Manusia Perubahan Serta Pengukurannya*. Balai Aksara, Yudistira dan Saadiyah. Ghalia Indonesia, Jakarta Timur. Cetakan kedua 1982.
- [21] Melage, Robert F. *Developing Attitude Toward Learning*. Penerjemah : Drs. A. Shofyanis, M.Ed & Thun Surjaman. Rosdakarya. Bandung, 1987.
- [22] Mustaqim & Abdul Wahib. 2010. *Psikologi Pendidikan*. Rineka Cipta, Jakarta.
- [23] Newcomb, Theodore M. et. al. 1981. *Psikologi social*. Terjemahan Bandung; CV. Dipenogoro.
- [24] Nunnally, Jum C. 1972, *Educational Measurement and Evaluation*, New York: McGraw-Hill Inc.
- [25] Rahayuningsih, Sri Utami. 2008. *Sikap (Attitude)*. Psikologi Umum.
- [26] Russefendi, E. T. 1989. *Pengajaran Matematika Modern untuk Orang Tua Murid, Guru, dan SPG*.
- [27] Silverius, Suke. 1991. *Evaluasi Hasil Belajar Umpan Balik*. Jakarta: PT Grasindo.
- [28] Slameto. 1991. *Proses Belajar mengajar Dalam Sistim Kredit Semester SKS*. Bumi Aksara, anggota IKAPI. Jakarta: Rineka Cipta.
- [29] Sternberg, Robert J. & Elena L. Grigorenko. 2009. *Mengajarkan Kecerdasan Sukses Meningkatkan Pembelajaran & Keberhasilan siswa*. Yogyakarta. Pustaka Pelajar.
- [30] Suherman, Erman, et. al. *Strategi Pembelajaran Matematika Kontemporer*. Common Text Book (edisi revisi). 2010
- [31] Surapranata, Sumarna. 2004. *Panduan Penulisan Tes Tertulis Implementasi Kurikulum 2004*. Bandung: Remaja Rosdakarya.
- [32] Suryabrata, Sumadi. 2008. *Psikologi Pendidikan*. Jakarta ; PT. Raja Grafindo Persada.
- [33] Tessmer, Martin. 1995. *Planning and Conducting Formatif Evaluation*. London: Biddles, Ltd, Guildford and King 's Lynn.
- [34] Tim PPs UHAMKA. *Model Proposal*, Program Pascasarjana UHAMKA
- [35] _____, *Pedoman Tesis dan Disertasi*, Program Pascasarjana UHAMKA
- [36] Tirtarahardja, Umar dan S. L. La Sulo. 2005. *Pengantar Pendidikan* : Jakarta PT. Rineka Cipta.
- [37] Tola, Burhanuddin. 2007. *Panduan Penulisan Soal Pilihan Ganda*. Jakarta: Pusat Penilaian dan Pendidikan Balitbang, Kemdiknas.
- [38] Tondowidjojo, JVS. 1993. *Kunci Sukses Pendidik*. Yogyakarta: Kanisius.

AN EVALUATION MODEL OF CHARACTER EDUCATION IN SENIOR HIGH SCHOOL

Hari Sugiharto¹, Djemari Mardapi²

¹Ministri of Education and Culture

²Yogyakarta State University
hari.soegiharto@yahoo.com

Abstract--This study aims to establish an evaluation model of character education that can provide information for school leaders and teachers of subject matters, both in terms of content, scope, format and timing of delivery, as well as beneficial for character education in senior high school. The subjects included the students and teachers who taught subject matters in the class. The instruments were validated with the construct validity technique using Confirmatory Factor Analysis (CFA) to test the suitability of the measurement models. Structural equation modeling (SEM) was used to analyze the suitability of the evaluation model. Both CFA and SEM used LISREL 8.80 to analyze the data. The suitability models, measurement model, and evaluation model, used the following criteria: 1) the value of load factor (λ) > 0.3, 2) the probability or p-value > 0.05, and 3) Root Mean Square Error of Approximation (RMSEA) < 0.05, Goodness of Fit Index (GFI) > 0.9. The results of the data analysis of this research can be summarized as follows. (1) The CEEM model is a good model for evaluating character education in senior high school, based on the experts' judgment, assessment by the users and the practitioners of character education, and the results of the analysis with LISREL 8.80 program, indicating a compatibility between hypothetical CEEM model with field data from the questionnaire (p-value = 0.140, GFI = 0.96 and RMSEA = 0.023), from the implementation for teachers (p-value = 0.056, GFI = 0.96, RMSEA = 0.025); from the actualization (p-value = 0.399, GFI = 0.98, RMSEA = 0.010); and from the implementation for students (p-value = 0.239, GFI = 0.99, RMSEA = 0.011). (2) The results of the experts' judgment, assessment by the users and practitioners show that the model evaluation guide of CEEM is better used as a reference in implementing the model; (3) The CEEM model is beneficial to determine the values, strengths, and weaknesses of character education intended to revise in order to increase the model's attractiveness and effectiveness.

Keywords: *model, evaluation, character education in Senior High School*

I. INTRODUCTION

Implementation of character education in high school is not functioning properly, this is evidenced by juvenile delinquency conducted particularly by learners at senior high level (SMA) is very worrying. Many brawls conducted by high school students is increasing and have caused fatalities. Data from the Indonesian Child Protection Commission (2012) mentions that there are 139 cases of brawl have occurred in Jakarta. Statistics from the Ministry of Health (2014) states that the sequence of cases of HIV/AIDS, Jakarta ranks third with 7477 cases, after Papua with 10.184 cases and 8976 cases in East Java. According to Kompas (Indonesian formal newspaper) (2013), in 2011, the number of high school students using drugs are about 3.187 people in 2012 and increasing into 3.410 people, and the new cases in 2013 recorded about 519 people. If there's no prevention effort then this figure will continue to rise.

Research of Williams and Mary (2000: 32) identifies issues of character education in schools include the following: (1) morality is a private matter and should be taught by families and places of worship, not in school, (2) the moral issue is very individual, so the school is impossible to teach it to students, (3) many educators do not have the competence to teach morals to students, (4) morality comes to us from the divine source that cannot be taught in a secular context, (5) the teaching of character education in schools would make religion as a part of school, (6) the time required to teach character sacrificing more important subject to learn.

Rawana; Franks; Brownlee; Neckoway (2011: 12-144) explains that one of the goals of character education curriculum is to bring the character to a higher level to be more in real life and better. Scott; Smith; Sunderland; Ward (1998) explains that the character values are the

goals that are attached to educational institutions and entered into a written curriculum. In addressing this, the Ministry of Education and Culture have a strategy in dealing with prevention of the decline of morals among the young generation. According to Main Design of Character Education (2011: 28), the explanation is that the implementation of character education at the school consists of three things. 1), the integration of character education in all learning materials (intra). 2), the integration of character education in extracurricular activities, and 3) integration through the school culture. But the strategy itself is apparently not enough.

The conclusion is that evaluation on optimal character education, there are number of factors should be taken into consideration in order to optimize the evaluation such as: 1) a part of the school is not optimal to evaluate the implementation of character education, 2) not all educators can be utilized as a model of implementation of the values of character, 3) some educators have not been optimal to insert instill character education in learning, 4) the integration of character values in the culture of the school has not gone well, and 5) the absence of evaluation model.

Problematic of primary schools face when evaluating their character education is not operational guidelines yet in evaluating the character education. School, until now has not had an evaluation model that is capable to evaluate the character education of learners appropriately, efficiently and effectively. With the expected school evaluation model that can capture information about the state of the character of the students at this time, so it can be repaired properly.

II. RESEARCH METHODS

Researchers using the Research and Development (R & D) method developed by Borg and Gall (2007: 775). The procedure is applied to the six-step development that are; 1) the study of theory and research results, 2) design evaluation model, 4) validation expert, 5) test products, and 6) the data analysis. Design model successfully arranged along with instruments and devices is an early draft of the model developed. After the evaluation model and its instruments and devices are developed, the next activity followed by validation of the experts (expert judgment) to construct and content of EPENKAR conducted in the form of FGD by including experts. 1). The experts in Education Evaluation, 2). Research Methodology, 3). Character Education, 4). Language, 5). School Culture, 6). Teachers, and 7). School Subject Teachers

A. Subject Test Research

Subject research trials are 31 high schools in Jakarta, with respondent subjects are teachers who taught at school and the students of class XI. The students of grade XI were chosen because these students already have got a character education for approximately 2 years.

B. Character Education Evaluation Model (Model EPENKAR)

Character Education Evaluation Model (EPENKAR) is a specific form of a set of components and procedures sequentially to realize an evaluation process of character education. A set of procedures is defined some procedures incorporated in a single unit. In figure 3, EPENKAR explained that the model consists of two major components, namely: 1) the implementation of character education that consists of: a) intra-curricular activities, b). extracurricular activities and c). school culture. 2) the actualization of the values of character education consisting of; a) process to think, b) process to feel/mind the heart, c) process to body, and d) process to initiative.

C. Data Collection and Data Analysis

The research instrument is a tool used by researchers to study data collection by conducting measurements, and written guidance in the form of a list of prepared questions to obtain information from respondents. By performing measurements, the obtained data was objective data needed to produce objective research conclusions. Data collection instruments required by the researchers consists of 6 devices, namely

1. Questionnaire on the implementation of character education (filled out by the teacher)

These instruments are used to determine the implementation of character education in schools that includes: 1). Intra-curricular, 2). extracurricular, and 3) culture of school. Data obtained using a questionnaire sheet in school. The data were obtained using Likert scale with 5 options.

2. The instruments of implementation of character education teacher (filled out by the teacher)

These instruments are used to study the behavior of teachers in schools of the values of character, namely: a) spiritual, b) honest, c) discipline, d) the responsibilities, e) tolerance, f) mutual help, g) manners, and h) confidence. The data were obtained using Likert scale with 5 options.

3. Assessment instruments of the implementation of character education to learners (filled by learners)

These instruments are used to study the behavior of learners in schools to the values which are: a) spiritual, b) honest, c) discipline, d) responsibilities, e) tolerance, f) mutual help, g) manners, and h) confidence. The data were obtained using Likert scale with 5 options.

4. Assessment instruments actualization of the values of character education to the learners (filled out by the teacher)

These instruments are used to assess the character of students, by observing the daily life of students in school. The data were obtained using Likert scale with 5 options.

D. Data analysis technique

The instrument of collecting data analyzed is by Confirmatory Factor Analysis (CFA), using the help of program LISREL 8.80 by using a second confirmatory order. Indicators are said to have good validity of the constructs or latent variables if: 1). t- value (factor loadings) > 1.96. (Doll, Xia, and Torkzadeh, 1994: 458) and 2). having Standardized Loadings Factor (SLF) at least ≥ 0.30 (Mooi & Sarstedt, 2011: 251), while Hair (2014: 102) states the terms of reliability is good is if it has a value of Construct Reliability ≥ 0.70 . As for the criteria on the size of the Goodness of Fit (GoF), namely: 1) p- value > 0.05, 2). Root Mean Square Error of Approximation (RMSEA) < 0.08, Goodness Fit Index (GFI) > 0.90. While supporter criteria Goodness of Fit (GoF) if NNFI, NFI, IFI and CFI, worth > 0.90, this means that overall it can be concluded that the model measured showed a good fit (good fit).

III. RESULT DEVELOPMENT

Character Education Evaluation Model (EPENKAR) in high school developed through three phases of testing, there are: 1) test of the construction and content, 2) test a limited scale, and 3) a broad scale trial. They are as follows:

1) test the construct and content

Before being tested, the instrument is validated by experts so that the instrument can perform its functions in accordance with the purpose of measuring the measurements were taken. Clarity construct research variables EPENKAR models validated in the form of Focus Group Discussion (FGD). Experts (Expert judgment) who participated in total are 9 people. The results of the validation using the techniques coefficient content validity (content validity coefficient) Aiken's V presented in Table 1.

Table 1. Summary of Results of Analysis of the Content Validity Aiken's

Validator	Jenis Instrumen		Σ butir	V terendah	Nilai Kritik
Expert Judgement	a.	Questionnaire	33	0,844	$V \geq 0,79$
	b.	Instruments for Teacher	50	0,844	$V \geq 0,79$
	c.	Instruments for Student	50	0,813	$V \geq 0,79$
	d.	Assessment instruments actualization	48	0,844	$V \geq 0,79$

Table 1 shows that all the instruments have a coefficient of Aiken's coefficient $V \geq 0.79$ which indicates that valid research instrument.

2) A limited scale trials

The trial results are limited, the coefficient of Cronbach Alpha reliability and the validity which are presented in Table 2, indicating that the instrument has a good reliability and validity.

Table 2. Summary of Results of Pilot Scale Limited

Jenis Instrumen	Σ item	Cronbach's Alpha	Corrected Item-Total Correlation $\geq 0,30$
Questionnaire	33	0,954	33
Instruments for Teacher	50	0,987	50
Instruments for Student	50	0,901	50
Assessment Instruments Actualization	47	0,980	47

3) The Wide-Scale Trial

a. Analysis of Factors Affecting Implementation Questionnaire Character Education

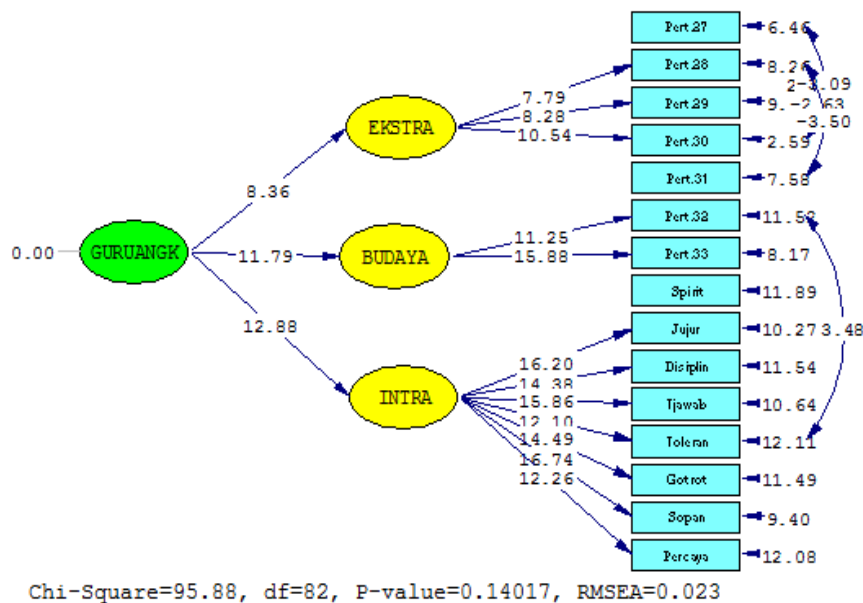


Figure 1. Variable CFA Teacher Questionnaire (t-value)

1. In figure 1 shows, construct questionnaires proved valid (p-value = 0,14017; GFI = 0.96; RMSEA = 0.023 and CFI = 1).
2. The instrument questionnaire has construct reliability (0.84 \geq 0.70) and t-value (intra = 12.88; extracurricular = 8.36; school culture = 11.79) $>$ 1.96. This shows that the reliability and validity of the model as a whole is good.

b. Analysis of Factors Affecting Implementation of Teacher

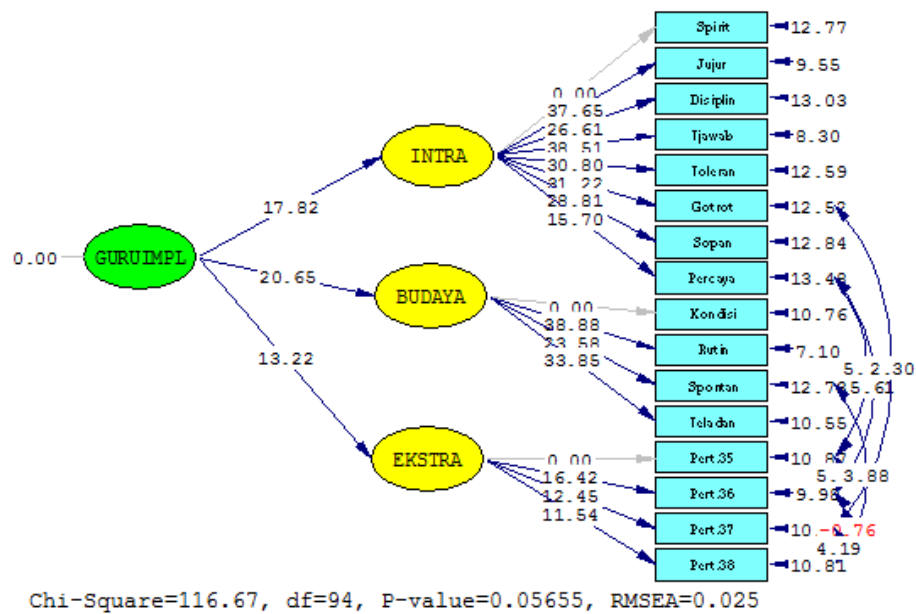


Figure 2. Implementation CFA Variable Master (t-value)

1. In Figure 2 shows, the construct of teacher implementation proved valid instruments (p- value = 0, 05 655; GFI = 0.96; RMSEA = 0.025; and CFI = 1).
2. The instruments of implementation for teachers to have the construct reliability ($0.93 \geq 0.70$) and t_{value} (intra = 17.82; extracurricular = 13.22; school culture = 20.65) > 1.96. This shows that the reliability and validity of the model as a whole is good.

c. Analysis of Factors Affecting Rate Actualization

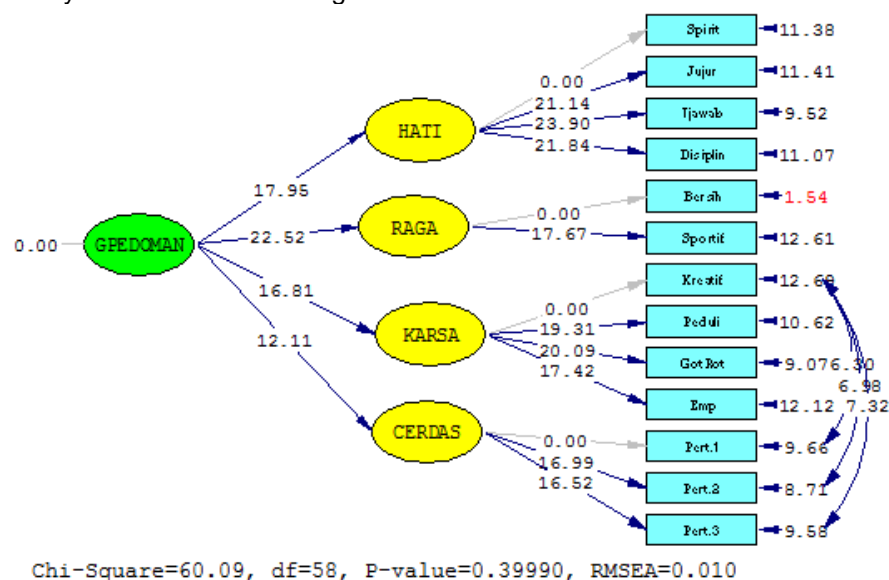
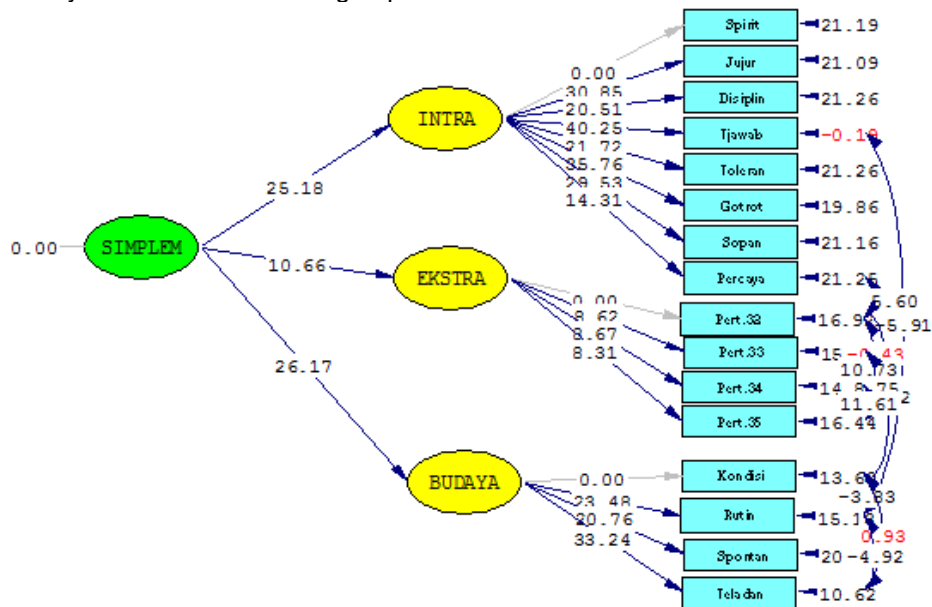


Figure 3. CFA Variable Teacher Guidelines (t-value)

1. In figure 3 shows, construct instruments actualization, proved valid (p-value = 0. 39 990; GFI = 0.98; RMSEA = 0.010; and CFI = 1).
2. Instruments actualization has construct reliability ($0.93 \geq 0.70$) and t value (process to feel with heart = 17.95; process to body = 22.52; process to initiative = 16.81, process to think = 12.11) > 1.96. This shows that the reliability and validity of the model as a whole is good.

d. Analysis of Factors Affecting Implementation of Students



Chi-Square=100.19, df=91, P-value=0.23933, RMSEA=0.011

Figure 4. Implementation Student in CFA variable (t-value)

1. In figure 4 shows, the student constructs proved valid instruments (p-value = 0.23933; the value of GFI = 0.99; RMSEA = 0.011; and the value of CFI = 1).
2. The instruments of implementation for the students to have the construct reliability ($0.92 \geq 0.70$) and t-value (intra = 25.18; extracurricular = 10.66; school culture = 26.17) > 1.96 . This shows that the reliability and validity of the model as a whole is good.

Evaluation of character education in high school as a model evaluation instrument is quite simple in execution, but the information revealed so complete, so it is an alternative that can be implemented by the leadership of the school to conduct an evaluation of the quality of character education in the school environment. Instruments character education evaluation model has been tested whose results showed that the implementation of this evaluation model is practical, economical, and objective.

Quantitative test results, for the student instrument analyzed by second order Confirmatory Factor Analysis (CFA) showed that the grains are instruments developed is valid and reliable. Free evaluation of character education suitable to evaluate the quality of character education in high school, because it is quite brief and incomplete, thus simplifying the implementation of character education evaluation system.

IV. CONCLUSIONS AND SUGGESTIONS

Based on the conclusions above, the following suggestions can be formulated:

1. Model EPENKAR can be used as an alternative for the Ministry of Education and Culture, Directorate of School, Department of State, Department of Education Office of the city/county and especially in schools, as the executor in the evaluation of character education in high school.
2. Model EPENKAR can be developed further in order to be more perfect for evaluation model of EPENKAR since it hasn't involved an independent appraiser (independent appraisal) from the outside. Therefore, in the future development should be considered to involve an independent appraiser, to integrate such assessments into the model of EPENKAR.
3. EPENKAR model is very simple that can be conducted by the user highly. This model can be further developed by a computer program, so that the evaluators can analyze data quickly and accurately.

REFERENCES

- [1] Benninga, Jacques., Wynne, Edward. (1998). *Keeping in character: A time-tested solution. Phi Delta Kappan*: Vol.79, Iss. 6; pg. 439, 8.
- [2] Benninga, Jacques., et.al, (2010). *Character and academics: what good schools do, Curriculum Leadership, readings for developing quality educational programs*. (9th ed.). United States of America: Pearson Education, Inc.
- [3] Borg and Gall. (1983). *Educational Research*, (4th ed.). Longman Inc., 95 Church street, white Plains, N.Y. 10601.
- [4] Doll, William., Xia, Weidong., Torkzadeh, G. (1994). A confirmatory factor analysis of the end user computing satisfaction instrument. *Management Information System Research Center. MIS Quarterly*, Vol. 18, no. 4 (dec. 1994), pp. 453-461
- [5] Hair, Joseph., et.al, (2014). *A Primer on partial least squares structural equation modeling (PLS-SEM)*. United States of America: SAGE Publication, Inc
- [6] [Huston-Holm, Pat](#). (2010). Effective character education at sentinel career center **techniques 85.6**, 8-9.
- [7] Kementerian Kesehatan (2014). Statistik Kasus HIV/AIDS di Indonesia. Diambil pada tanggal 3 Januari 2015, dari <http://spiritia.or.id/Stats/StatCurr.pdf>.
- [8] Komisi Perlindungan Anak Indonesia. (2012). *Tawuran pelajar 2012*. Diambil tanggal 2 Januari 2015, dari <http://fandibree13.blogspot.com>.
- [9] Kompas. (2013). Pengguna narkoba di kalangan remaja meningkat. regional. Diambil pada tanggal 3 Januari 2015, dari <http://www.kompas.com>
- [10] [Milson, Andrew J.](#) (2002). *Introduction: developing a comprehensive approach to character education in the social studies*, **the social Studies 93.3**, 101.
- [11] Mooi, Erik., Sarstedt, Marko (2011). *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistic*. Springer-Verlag Berlin Heidelberg.
- [12] Parkay, Forrest., Anctil, Eric & Hass, Glan. (2010:6). *Curriculum Leadership, readings for developing quality educational programs*. (9th ed). Boston: Pearson.
- [13] [Rawana, Justin.](#) et al,(2011).The application of a strength-based approach of students' behaviours to the development of a character education curriculum for elementary and secondary schools. **The Journal of Educational Thought,45.2**: 127-14.
- [14] [Williams, Mary M.](#) (2000). Models of character education: perspectives and developmental issues. **Journal Of Humanistic Counseling, Education And Development 39,1**: 32.

AN EVALUATION ON THE IMPLEMENTATION OF LESSON PLANS FOR EARLY CHILDHOOD EDUCATION CENTER (PAUD) LOCATED AROUND IAIN SURAKARTA

Hery Setiyatna

Islamic Kindergarten Study Program, Education and Teacher Training Faculty, State Islamic Institute of Surakarta (PGRA FITK IAIN Surakarta);
Indonesian Education Evaluation Association in Regional Work Unit of Central Java (HEPI UKD Jawa Tengah)
herysetiya@gmail.com

Abstract— The paper aims to describe the implementation of lesson plan for Early Childhood Education (PAUD) located around IAIN Surakarta. This is the study on evaluation. 23 PAUD centers were taken as samples. The PAUD Principals are the informants for this study. The data were collected through observation, interview, and documentation. Data analysis was done through percentage using descriptive qualitative.

Three findings are; (1) most of PAUD centers implemented School-based curriculum (KTSP) while some of them still implemented K-13 curriculum with some considerations. (2) each PAUD center has the completed lesson plan (3) the process of curriculum development on each PAUD center has the different ways (a) teachers' discussion together with PAUD principal (b) individual discussion among the teacher then checked by the PAUD principal (c) teachers' discussion through several forums (IGTK, IGRA, IGBA dan IGAUD) and make an agreement.

Keywords: *Implementation, Lesson Plans, and TK/RA/BA/PAUD*

I. INTRODUCTION

Learning system is composed in systemic and systematic way. It has four areas. They are (1) planning, (2) implementing, (3) assessing and (4) reflecting. Planning is the phase in which teachers are required to compile the lesson plan. It is based on the curriculum that is the base of Educational planning. as a reference for an educational program. Curriculum as an educational strategy is designed to achieve national education goals. The Government of indonesia under The Ministry of Education and Culture realizes The importance of the curriculum and its role to give directions for the educational program to establish the output of National Education.

Early Childhood Education (PAUD) is the most fundamental education for the child's development that is determined by a variety of meaningful stimulation given from an early age. PAUD must be prepared in a planned and holistic-integrative so that a golden period of the development of the child-stimulated intact, to develop their potential. To achieve that goal, a structured education program should be implemented. Components for structured education is the curriculum. Early Childhood Education intended in this study is the educational institution in society commonly referred to as kindergarten (TK) / Islamic kindergarten (RA/ BA) / Early Childhood Education Center (PAUD).

The curriculum guides teachers to meet all learning areas used by children using the effective learning approach, the strategy of structuring a suitable environment to support the development of the quality of children in accordance with their development stage. Teachers must have a strong vision to develop their children, knowledge and skills that children would learn the attitudes and values so that they will apply them in their daily lives. Children need appropriate direction and guidance from the teachers.

PAUD curriculum is expected as the base for the preparation of children in order to be better children. The goal is to prepare them for reaching their higher education. Delivering early childhood, which is ready to continue education is not only limited to the child's ability in reading,

writing and arithmetic skills but in all aspects of child's development. In 2014/2015 academic year, the 2013 National Curriculum (K-13), is implemented. The K-13 is as the guideline to replace the previous curriculum (The School-Based Curriculum/ KTSP) which has (1) dimensional plans concerning the objectives, content and learning materials, (2) techniques used for learning activities [2].

Learning is a real implementation of education in the educational unit. Good learning is presented systematically and continuously. The learning activities are designed to follow the principles of learning, to consider the suitable materials, learning experiences, allotted time and setting, tools / learning resources, classroom organization and method of evaluation. In PAUD, the activities implemented through meaningful games to get children's learning experiences.

The quality of learning is measured by changes in the behavior of children in the appropriate direction referring to the based competence. Therefore, PAUD teachers are expected to design, to develop and to implement learning activities that correspond to the characteristics and the needs of children.

Lesson plan is designed for teachers to implement activities that facilitate children in the learning process. Lesson plan should be arranged before the learning activities undertaken. The role of the teachers are as a conceptor/manager, so that the lesson plan must be prepared by the teachers independently. There are three types of lesson plan should be prepared by the teacher before implementing learning process; (a) Programs on each Semester (ProSem), (b) Weekly lesson plan (RPPM) , and (c) Daily Lesson Plan (RPPH). The teachers' knowledge to create good lesson plan as continuous task is an interesting topic to study. The aim of the study is to describe the implementation of the teaching and learning in several PAUDs located around IAIN Surakarta.

This study aimed to describe the implementation of lesson plans for PAUD located around IAIN Surakarta. The study will focus on (1) the curriculum (2) the lesson plans (3) the process to develop the lesson plans.

II. METHODOLOGY

Based on the type of data and research objectives, the study was categorized as an evaluation research using qualitative design to study the meaning conveyed by the respondents about problems or issues research [1] [6]. The research was conducted in April 2016 and took 23 early childhood centers/PAUD around IAIN Surakarta which consist of TK, RA, BA and PAUD from several cities. 23 PAUDs were taken as samples. They were one center in Surakarta (4.4%), 3 centers in Karanganyar (13.0%), 5 centers in Boyolali (21.7%), 2 centers in Klaten (8.7%), and 12 centers in Sukoharjo (52.2%). The reasons of selecting Those cities is because many community service activities of the lecturers of IAIN Surakarta are conducted in those cities. The researcher also wanted to know the capability of the teacher to make lesson plans. The subjects of this study consisted of 23 PAUD teachers around IAIN Surakarta. The Informants in this study was the PAUD principal around IAIN Surakarta. The data collection was done through observation, interviews and documentation. Data analysis was presented by percentages and descriptive qualitative [5].

III. FINDINGS AND DISCUSSION

A. Findings

The Results of the study on The implementation of PAUD lesson plan around IAIN Surakarta are as follows: (1) the curriculum (2) the process to plan the lesson. (3) the process to develop the lesson plans.

The researcher found that those PAUDs implemented two models of curriculum as the basis for the lesson plan. Those models are the School-Based Curriculum/ KTSP and 2013 curriculum/K-13. The data showed that, 15 PAUD centers (65,3%) implemented the School-Based Curriculum/KTSP and 8 PAUD centers (37,7%) implemented K-13 curriculum. Completeness of lesson plans prepared at PAUD center was found that all PAUD centers (100%) had a complete lesson plan, they are (a) Annual Program, (b) Semester Program (c) Weekly Program (Weekly Lesson Plan/RPPM) and (d) Daily Program (Daily Lesson Plan/RPPH).

Meanwhile, The process of planning the development of learning at PAUD center around IAIN Surakarta was found in three ways, they are (a) the process was done through discussion led by the principal of PAUD conducted by seven PAUD centers (30.4%), (b) the process was

done through assigning teachers individually, then the results were checked by the PAUD principal and organized by five PAUD centers (21.8%), (c) teachers discuss through the organization (IGTK, IGRA, IGBA and IGPAUD) the results of the discussion applied by participants in the discussion and organized by 11 PAUD (47,8%).

B. Discussion

The curriculum implemented as the basis for planning the implementation of PAUD was School-based curriculum/KTSP despite the facts that curriculum K-13 PAUD has been released since 2014/2015. It is because the teachers still waiting for "the instruction" from the head of educational district, school supervisors and even from the PAUD center. Although some PAUD centers have prepared the training on the implementation of the K-13 and had prepared a lesson plan based on K-13, but they still implement the school-based curriculum because the PAUD center has not received instructions from superiors or the PAUD management. Teachers Personally feel unprepared to implement the K-13 curriculum for PAUD, because of limited information time to inform it. it is suggested that the pre-implementation process of K-13 curriculum for PAUD should be followed by monitoring and evaluating to PAUD center in the form of assissting. Some PAUD principal implementing KTSP curriculum stated that the PAUD properties were still not satisfied. This is one of the factors that make the PAUD centers were not ready to implement K-13 curriculum. In the era of Ministry of educational affairs Anies Baswedan, K-13 is now still being revised by the government, so that the PAUD centers prefer implementing KTSP to K-13.

PAUD centers that already implement the lesson plan using K-13 aim to foster children love to learn, happy to do the scientific experiments, not to emphasize on the mastering the subjects or assessing the process on early childhood referring to their developmental stages. The learning process which is done through games and activities also provide additional knowledge, attitudes, and skills of the child in accordance with the Basic Competence using appropriate attention to the capacity developmental stages of children at certain ages in general [4].

The complete lesson plan for PAUD consists of (1) Program Semester, which contains (a) List of themes that will be developed into sub-subthemes, (b) Competence that is set to be achieved on each theme and time allocation of each theme. To determine the Basic Competence on each theme, six development programs are included, Religion and Moral Values, Motoric, Cognitive, Social &Emotional, Language and Arts. The Format of Semester Program plan is done by PAUD center. (2) RPPM is composed for the weekly lesson plan. RPPM is derived from Semester Program and contains specific plans. It includes (a) the service program identity, (b) selected basic competence, (c) learning materials and (d) activity plans. (3) RPPH, is is composed for the daily lesson plan. RPPH is formulated and implemented by educators. There is no standardized format of RPPH, but it contains components that are established, it includes (a) Program Identity, (b) materials, (c) Equipment and Materials, (d) Opening activity, (e) Main Activities, (f) Closing activity, and (g) Assessment Plan.

Teachers have main responsibility to plan and to run the curriculum for PAUD. In practice, the effort of the planning the Semester Program, RPPM and RPPH is implemented in various ways. There are teachers who develop RPPH after completing the learning process and develop RPPM carried out together once in a week (every Thursday or Saturday). There is PAUD agency that develops Semester Program, RPPM and RPPH on the previous semester but it does not interfere with the learning process. Association of Islamic Kindergarten (IGRA) conducts a meeting to discuss the Semester program, RPPM and RPPH. The monthly meeting is held regularly, at an initial meeting is to develop lesson plans, after the program is agreed, at a meeting in the next month they have discussion on the learning process that occurs in each PAUD center based on the agreed program. The advantages and disadvantages are discussed together, if necessary, they will revise the lesson plan.

The Annual and Semester program of PAUD centers are conducted by the city government under the Educational affairs section in collaboration with Religious Affairs section along with PAUD management/foundation. The Development of weekly Action Plan (RKM) or RPPM is based on the Annual Program and Semester Program originating from Educational affairs section in collaboration with Religious Affairs section, but the RKM in this PAUD center is always the same and static. It is from the year 2011 - now.

IV. CONCLUSION

Based on the discussion above, the researcher draws the conclusion as follows: (1) 65.3% of PAUD center still implement school-based curriculum/ KTSP as the base to make lesson plan (2) Completeness of lesson plan has been complete. It comprised of (a) Annual Program, (b) Semester Program, (c) Weekly Program: RKM or RPPM, and (d) Daily Programme: RKH or RPPH. (3) The process of lesson plan implementation in PAUD around the campus IAIN Surakarta reached 47.8% assigning teachers to discuss with the organization (IGTK, IGRA, IGBA and IGPAUD). The results can be applied to other samples of PAUD centers.

REFERENCES

- [1] Creswell, J.W., Research Design: Qualitative, Quantitative, and Mixed Method Approaches. Los Angeles: Sage Publications, Inc., 2009.
- [2] Direktorat Pembinaan PAUD, Dirjen PAUD dan PenMas, Kemendikbud, Kerangka Dasar dan Struktur Kurikulum 2013 PAUD. Jakarta: Tidak diterbitkan, 2015.
- [3] Direktorat Pembinaan PAUD, Dirjen PAUD dan PenMas, Kemendikbud, Pedoman Penyusunan Rencana Pelaksanaan Pembelajaran PAUD. Jakarta: Tidak diterbitkan, 2015.
- [4] Luluk Asmawati. Perencanaan Pembelajaran PAUD. Bandung: Remaja Rosdakarya. 2014.
- [5] Patton, MQ., Metode Evaluasi Kualitatif, Penerjemah: Budi Puspa Priyadi. Yogyakarta: Pustaka Pelajar. 2009.
- [6] Wilig, C., Introducing Qualitative Research in Psychology. Glasgow: Bell and Bain, Ltd., 2008.

THE EFFECT OF COOPERATIVE LEARNING MODEL TYPE GROUP INVESTIGATION WITH SELF ASSESSMENT REINFORCEMENT AND LEARNING INTEREST TOWARD THE PHYSICS LEARNING RESULT OF STUDENTS AT GRADE XI SMA NEGERI 1 WATUBANGGA KOLAKA

I gede Purwana Edi Saputra¹, H.M. Sidin Ali,²

¹Student at Phisycs Education Post Graduate Program, State University of Makassar

²Lecturer at Phisycs Education Post Graduate Program, State University of Makassar
gedepurwana@gmail.com

Abstract—This study objectives to determine differences the physics learning results among students are taught by cooperative learning model type group investigation with self assessment reinforcement (GISA) and students who are taught by conventional model from the aspect of learning interest, and to determine the interaction between the learning model and learning interest toward learning results. This study is a quasi-experimental with 2x2 factorial design. The population includes all students at grade XI.IPA SMA N 1 Watubangga consisting of 5 classes. The samples is taken by random of 4 class that is 2 classes as the experimental group and 2 classes as the control group. The results showed that: (1) Overall, there are differences in physics learning results between students are taught by GISA and students who are taught conventionally; (2) for a group of students who have a high learning interest, there are differences in learning results between taught by GISA and that taught by conventional; (3) for a group of students who have low learning interest, there are differences in learning results between taught by GISA and that taught by conventional; (4) There is no interaction between the learning model and learning interest toward the physics learning results of students.

Keywords: *GISA, conventional, learning interest, learning results*

I. INTRODUCTION

Along with changes the paradigm of learning, the success of teaching and learning in an education, in that it is a school, both elementary and intermediate school is not only determined by teacher, but was strongly influenced by active students. In studying physics most of us are dealing with concepts and various laws that arrange the concept. This means thinking processes of student should be integrated into learning from things that are concrete to the abstract so that in the process of teaching and learning physics the student must continuously find and develop the concept of regular and planned so as to understand the concepts physics well. These characteristics have implications on how physics is taught so that students are able to understand the physics integrally. What kind of learning activity that is expected by students and teachers, resulting in process quality as well as maximum learning result? The question is difficult to answer by most teachers or may be answered but never apply them in learning activities in the classroom. In general, the learning process is expected by students and teachers already presented to the Minister of National Education (Permendiknas) Number 41 of 2007. In Permendiknas was stated that the learning process for each subject must be flexible, varied, and fulfill the standards.

Learning physics is not only related to physical characteristics, but also should be adapted to the level of cognitive development of students. According to Piaget, cognitive development of junior-high school students is at the stage of formal operations, as it has aged 11 years and over. The learning activities in high school physics should provide the opportunity for students to optimize the ability of its formal operational thinking as reflective thinking and hypothetical-deductive^[10]. In practice, the teaching and learning process especially physics discovered many problems such as lack of motivation and interest students, and not optimal yet the capability of students in understanding and resolving the problems of physics that ultimately led to the low

performance of learning result of students. This problem was found in one of senior high school (SMA) in Indonesia, Southeast Sulawesi Province, namely in SMA Negeri 1 Watubangga. Based on observations conducted, there was information that the learning conditions in SMA Negeri 1 Watubangga especially in physics is as follows: (1) The process of learning that takes place is still dominated by learning one directional or focused on teachers; (2) Only a few students active in the learning; (3) The number of students who pass the daily tests subjects of physics is still not optimal.

The above facts indicate that the teacher has not been optimally understood the characteristics of physics and cognitive developmental characteristics of students, so that the learning process has not led to the activity of thinking students. One of the learning that is expected to increase the role of students in the learning process is a cooperative learning model type group investigation. This is because in a cooperative learning model group investigation largely the learning activities centered of students by providing greater opportunities for students to be creative and mutually express their opinions in order to solve the problems given. Cooperative learning model is seen as a process of active learning, because students will be more to learn through the process of formation (constructing) and creation, work in groups and share knowledge as well as individual responsibility remains the key to successful learning^[8].

If many researches have shown the effectiveness of cooperative learning model in improving the quality of learning physics, then how a teacher can ensure that students actually have done the components of cooperative activities? Or whether the cooperative process has been accomplished in the lesson? During this observational study materialize only done by the teacher, so that efforts to improve the learning of course only are done by the teacher. If so, then the students do learning activities not because of internal motivation, but because of the influence of outside intervention in this case encouragement from teachers. Learning will run naturally if the teacher is able to raise awareness of the importance of the material to be learned, and continue to encourage the emergence of self-regulation students. One way that can be done by teachers is to train students to self-assessment. Tierney states that self-assessment is part of the learning process, which can help students focus on what kind of work they have to do and show that they become more responsible for the direction of their learning^[6]. Andrade, Basnet, Nbina, Mansyur, and Ross in his research found that the implementation of self-assessment in learning a positive impact on learning result of students^{[1][2][4][5][7]}.

One thing that is important and needs attention in learning is the interest of students. The interest spawned spontaneous attention that allows the creation of concentration for a long time thus, interest is the foundation for concentration^[3]. According to interest as one of the psychological aspect is influenced by several factors, both the inside (internal) or outside (external). Someone who is interested in a particular subject, usually tend to pay attention to these subjects^[9]. Conversely, if a person is continuously paying attention either consciously or not a particular object, can usually generate interest in the object. Interests of students greatly affect how students build a strong stimulus in the learning process. Obviously this will impact on learning result of students. If the interest of learning support in the learning process of formation stimulation, of learning result will be good. Otherwise, the learning results to be not optimal. That is why the learning interest and assessment in early learning becomes an important tool to support learning. Thus the rationale encourage researchers to conduct research on the effect of cooperative learning model type group investigation with the self assessment reinforcement and interest in learning toward the physics learning result of students.

The problems in this research are: (1) Overall, whether there are differences in learning result between the physics students taught by cooperative learning model type group investigation by self assessment reinforcement (GISA) with students who are taught conventionally? (2) For a group of students who have a high learning interest, whether there are differences in learning result between the physics students are taught through learning model GISA with that taught by conventional? (3) For a group of students who have a low learning interest, whether there are differences in learning result between the physics students are taught through learning model GISA with that taught by conventional teaching? (4) Is there an interaction effect between learning models and interest in learning the physics learning result of students? The purpose of this study was to answer the problem is to determine differences in the results of learning physics among students taught through cooperative learning model type group investigation by self assessment reinforcement (GISA) with students who are taught by conventional learning both overall and in terms of aspects learning interest, and to understand the interaction between the

learning model and interest in learning toward learning result. The results of this study are expected to provide good benefits for students, teachers, schools and other researchers. For students, this study may provide a new experience in learning, especially with regard to self-assessment. For teachers this research will shed light on how the characteristics of the subjects of Physics and how to teach physics right through the reflection process by providing opportunities for students to self-assessment. For schools, the results of this study can be used as material / basic consideration in the preparation of school development plans are oriented to the needs and teachers, especially in the repair process of learning. As for other researchers, this study can be used as a reference to conduct further research on the self-assessment.

II. RESEARCH METHOD

This study used a sample of two groups: the experimental group and the control group. The experimental group was treated GISA learning model, while the control group with conventional learning. Each group is distinguished into two parts based on the learning interests of students, the student with high learning interest and student with low learning interest. This research was quasi experiment with a 2 x 2 factorial design as shown in Table 1 below.

Table 1. Research Design Of Factorial 2x2

Learning Interest	Learning Model		Σ
	GISA (A ₁)	KV (A ₂)	
High (B ₁)	A ₁ B ₁	A ₂ B ₁	A ₁ B ₁ + A ₂ B ₁
Low (B ₂)	A ₁ B ₂	A ₂ B ₂	A ₁ B ₂ + A ₂ B ₂
Σ	A ₁ B ₁ + A ₁ B ₂	A ₂ B ₁ + A ₂ B ₂	

The population in this study was all students grade XI in SMA Negeri 1 Watubangga, Kolaka, Southeast Sulawesi. Distribution of students in each class is shown in Table 2 below.

Table 2. Distribution Of Students Class XI IPA Sma Negeri 1 Watubangga

Class name	Number of student
XI.IPA-A	34
XI.IPA-B	36
XI.IPA-C	35
XI.IPA-D	35
XI.IPA-E	36
Total	176

The samples in this study are the four classes were randomly selected from five classes above using sampling random techniques (simple random sampling). As for the distribution of more samples in this study are presented in Table 3 below.

Table 3. Distribution Of Research Sample

Number	Sample	Class	Total
1	Experiment	XI IPA A XI IPA B	70
2	Control	XI IPA C XI IPA D	70
Total			140

The instrument used in this research are treatment instrument and measuring instruments. Treatment instrument in this research is a learning device in the form of Lesson Plan and Worksheets Student. The measuring instrument consists of a questionnaire students' interest in learning and test of cognitive learning result. To determine the distribution of interest category used to study physics division based normal distribution curve is selected 27% of the group of students to study physics learning interest high and 27% of a group of students to study physics lower interest.

Before the instrument is ready to use, it must be validated that is validation of content and empirical validity. Test validation of contents was conducted on the treatment and measurement

instruments. Analysis of the contents of the instrument validation is done in this research that uses Gregory models in the form of a model agreement among experts. Empirical validation test is only done to test the cognitive learning. This test aims to know the validity of item instrument and reliability of the instrument based on empirical data obtained through testing the. Types of data and data collection techniques used in this research can be seen in the following table.

Table 4. Types and Data Collection Techniques

Number	Data	Data types	Data Collection Techniques	Data Source	Instrument
1	Learning interest	Interval	Questionnaire	student	Interest learning questionnaire
2	Cognitive learning result	Interval	Writing test	student	Test item of cognitive learning result

Analysis of the data used in this study consisted of descriptive statistical analysis, inferential analysis prerequisite test and to test the hypothesis.

A. Normality test

Normality test is used to determine whether the data comes from populations studied normal distribution or not. Statistical test used was chi-squared (χ^2).

B. Homogeneity test

This test is used to determine whether the research data has a homogeneous variance. Homogeneity of the data used to calculate by the F- test.

C. Descriptive statistical analysis

Descriptive analysis functions are to provide an overview of the data obtained, such as: the number, maximum, minimum, mean, mode, median, standard deviation and variance.

D. Hypothesis Testing

Statistical analysis was used to test the main hypothesis of this research is the analysis of Two Way ANOVA, which is to measure the difference in cognitive learning result due to differences in the treatment of learning and the level of interest of students, as well as whether there is interaction both on learning result. Further test used in this study are t-test and Tukey-test.

III. RESULT AND DISCUSSION

A. The Result Of Data Analysis Research

1. Descriptive Statistic

Table 5. Score Statistic Learning Interest And Physics Learning Result

Learning result		Learning result		Learning result test		Number of student	
		GISA	Conventional	GISA	Conventional	GISA	Conventional
High interest	Min	113	103	18	15	19	19
	Max	131	130	23	21		
Low interest	Min	49	48	11	9	19	19
	Max	95	70	18	16		
Total						38	38

The results of descriptive analysis towards the physics learning result of student class experiment and control can be seen in Table 6.

Table 6. Score Recapitulation Learning Result Of Student Experiment And Control Group

Descriptive of Learning Result	Experiment Group	Control Group
Total	672	586
Maximum	23.00	21.00
Minimum	11.00	9.00
Mean	17.68	15.42
Modus	19.00	13.00
Median	18.00	15.00
Deviation standard	3.19	3.30
Variance	10.16	10.89

Here is a table of frequency distribution physics learning result test scores of students taught by GISA Model.

Table 7. Frequency Of Physics Learning Result Student Are Taught By GISA Model

Number	Interval Class	Frequency of GISA Model	Percentage	Category
1	1 - 5	0	0.00 %	Very Low
2	6 - 10	0	0.00 %	Low
3	11 - 15	9	12.85 %	Medium
4	16 - 20	45	64.29%	High
5	21 - 26	16	22.86%	Very High
Total		70	100.00%	

For more details, Table 7 is shown in histogram form as follows.

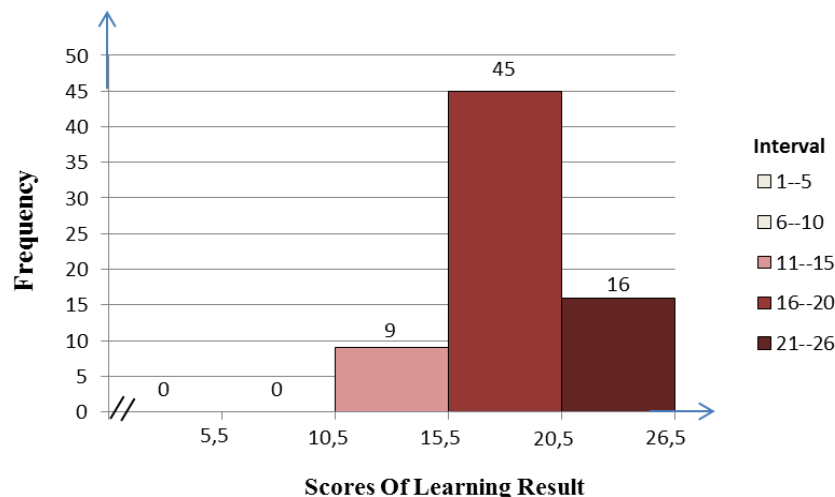


Figure 1. Histogram of Test Scores Physics learning result are taught by GISA Learning Model

The frequency distribution of physics achievement test scores of students taught using conventional learning model can be seen in table 8 below.

Table 8. Frequensi of Physics learning result student are taught using Conventional Model

Number	Interval Class	Conventional Model Frequency	Percentage	Category
1	1 - 5	0	0.00%	Very Low
2	6 - 10	3	4.29%	Low
3	11 - 15	24	34.29%	Medium
4	16 - 20	40	57.14%	High
5	21 - 26	3	4.29%	Very High
Total		70	100.00%	

For more details, the above table is displayed in histogram form as follows.

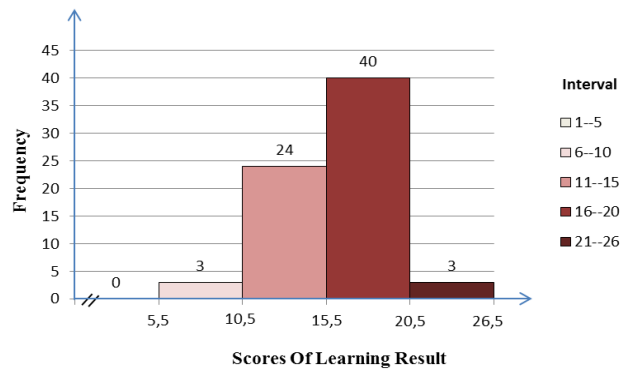


Figure 2. Histogram of Test Scores Physics learning result are taught by Conventional (KV) Model

2. Inferential Statistic

Before the test the hypothesis, first tested the normality and homogeneity of variance test data as a precondition test inferential statistics using parametric tests.

a) Test Of Normality

Table 9. The Result Of Normality Test

Class	Aspects Tested	Number of sample	$\chi^2_{\text{arithmetic}}$	χ^2_{table}	Information
Experiment	High interest	38	1.91	7.81	Normal
	Low interest				
Control	High interest	38	3.44	7.81	Normal
	Low interest				

b) Test Of Homogeneity

Table 10. Homogeneity of Variance Data

Testing Homogeneity of Variance				
Number	Group	Sample	dk	Variance (S^2)
1.	Experiment	38	37	10.16
2.	Control	38	37	10.89

The price of the F arithmetic compared with F table for significance 5% = 1,72 in order to obtain the price $F_{\text{arithmetic}} < F_{\text{table}}$ (1.07 < 1.72). It can be concluded that the variance data to be analyzed homogeneous so that Two Way ANOVA for hypothesis testing can proceed. To facilitate testing of the hypothesis then created a spreadsheet Two-Way ANOVA with the same cell as follows.

Table 11. Basic Statistic For ANOVA Test

Learning interest (B)	Learning Model (A)		Total ($\sum B$)
	GISA (A_1)	Conventional (A_2)	
B1	n = 19	n = 19	n = 38
	$\sum (x) = 385$	$\sum (x) = 342$	$\sum (x) = 727$
	$\sum (x)^2 = 148225$	$\sum (x)^2 = 116964$	$\sum (x)^2 = 528529$
	$\bar{x} = 20.26$	$\bar{x} = 18.00$	$\bar{x} = 19.13$
	$\sum (x^2) = 7843$	$\sum (x^2) = 6232$	$\sum (x^2) = 14075$
	S = 1.52	S = 2.05	S = 2.12
	$S^2 = 2.13$	$S^2 = 4.22$	$S^2 = 4.49$

B2	$n = 19$ $\sum (x) = 287$ $\sum (x)^2 = 82369$ $\bar{x} = 15.11$ $\sum (x^2) = 4417$ $S = 2.13$ $S^2 = 4.54$	$n = 19$ $\sum (x) = 244$ $\sum (x)^2 = 59536$ $\bar{x} = 12.84$ $\sum (x^2) = 3208$ $S = 2.03$ $S^2 = 4.14$	$n = 38$ $\sum (x) = 531$ $\sum (x)^2 = 281961$ $\bar{x} = 13.97$ $\sum (x^2) = 7625$ $S = 2.35$ $S^2 = 5.53$
Total ($\sum K$)	$n=38$ $\sum (x)=672$ $\sum (x)^2 = 451584$ $\bar{x} = 17.68$ $S = 3.19$ $S^2 = 10.16$ $\sum (x^2) = 12260$	$n=38$ $\sum (x)= 586$ $\sum (x)^2 = 343396$ $\bar{x}=15.42$ $S = 3.30$ $S^2 = 10.89$ $\sum (x^2) = 9440$	$n_T=76$ $\sum x_T = 1258$ $\sum (x_T)^2 = 1582564$ $\bar{x}_T = 17.607$ $\sum (x_T^2) = 21700$

We can see that the students who have high and low interest learning taught by GISA learning model has a total score of higher learning result than those taught by conventional learning model. Summary results of two-ways ANOVA can be shown in table 12.

Table 12. Result Summary of Two-Ways ANOVA

Source of variance	Df	Sum Of Square	Mean Of Square (s^2)	F_a	F_t	Testing Decision
Between Columns(Ak)	1	97.31	97.31	25.60	3.96	H_0 rejected
Between Rows (Ab)	1	505.47	505.47	133.01	3.96	H_0 rejected
Interaction (I)	1	0.01	0.01	0.002	3.96	H_0 accepted
Between Groups (A)	3	602.79	200.93	52.87	2.76	H_0 rejected
Within Groups (D)	72	274.00	3.80	-	-	-
Corrected Total (TR)	75	876.79	11.69	-	-	-
Intercept (R)	1	20823.21	20823.21	-	-	-
Total (T)	76	21700	-	-	-	-

B. Discussion

- Overall, there are differences in cognitive learning result of students' experimental group and the control group.

For the first hypothesis of the results of inferential statistical analysis using Two ways ANOVA can be seen the value of $F_{\text{arithmetic}}$ at the source of variance between columns, the price obtained $F_{\text{arithmetic}} = 25.60$. The value $F_{\text{arithmetic}}$ is then compared with the price F_{table} , for significance level $\alpha = 0.05$ with db = 1 obtained price $F_{\text{table}} = 3.96$. Because $F_{\text{arithmetic}} > F_{\text{table}}$ then H_0 rejected and H_1 accepted. This means that there are differences in physics learning result among students who were taught by GISA models and students are taught by conventional learning models.

- There are differences in cognitive learning result based learning model application on a group of students with high learning interest

For the second hypothesis of the results of inferential statistical analysis using Two ways ANOVA can be seen the value of $F_{\text{arithmetic}}$ at the source of variance between groups, obtained by the price of $F_{\text{arithmetic}} = 52.87$. $F_{\text{arithmetic}}$ value is then compared with the price F_{table} , to the significance level $\alpha = 0.05$ db = 3 price obtained $F_{\text{table}} = 2.76$. Because of $F_{\text{arithmetic}} > F_{\text{table}}$ then H_0 rejected and H_1 accepted. This means that for students with a high interest in learning physics there are differences in learning result among students who were taught by GISA Models and students are taught by conventional models.

- There are differences in cognitive learning result based learning model application on a group of students with low learning interest.

For the third hypothesis of the results of inferential statistical analysis using Two ways ANOVA can be seen the value of $F_{\text{arithmetic}}$ at the source of variance between groups, obtained by the price of $F_{\text{arithmetic}} = 52.87$. $F_{\text{arithmetic}}$ value is then compared with the price F_{table} , to the significance level $\alpha = 0.05$ db = 3 price obtained $F_{\text{table}} = 2.76$. Because of $F_{\text{arithmetic}} > F_{\text{table}}$ then

H_0 rejected and H_1 accepted. This means that for students with a high interest in learning physics there are differences in learning result among students who were taught by GISA model and students are taught by conventional learning models.

4. There is no interaction effect between learning models and interest in learning the cognitive achievement of students.

For the fourth hypothesis of the results of inferential statistical analysis using Two ways ANOVA can be seen the value of $F_{\text{arithmetic}}$ at the source of interaction variance, obtained the price of $F_{\text{arithmetic}} = 0.002$. Fhitung value is then compared with the price F_{table} , to the significance level $\alpha = 0.05$ db = 1 price obtained $F_{\text{table}} = 3.96$. Because of $F_{\text{arithmetic}} < F_{\text{table}}$ then H_0 accepted and H_1 rejected. This means that there is no interaction between applied learning and learning interest toward learning result.

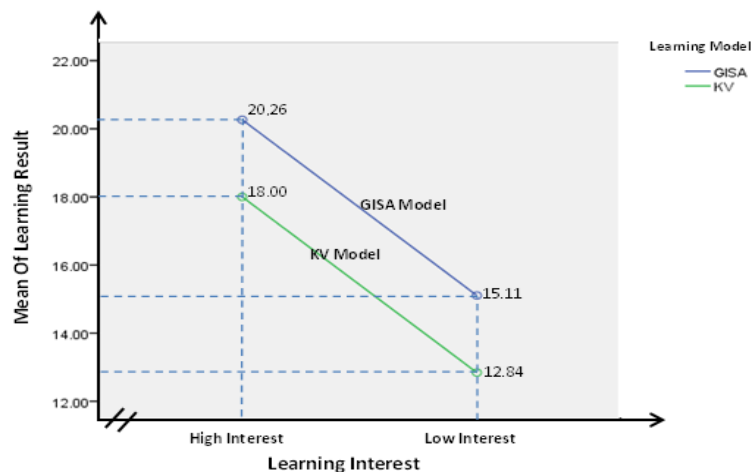


Figure 3. Graph interactions between Learning Model with learning Interest toward Learning Results

The absence of interaction in learning models and the learning interest toward learning result of this study show a treatment effect directly on classroom experiments using model GISA. This treatment effects in the form of reinforcement through self-assessment given to students every learning process. With the reinforcement given the growing Assurance factor (confidence) in good students who have a high interest in learning and ask for a low learning so that students are motivated to obtain better results in the achievement of learning result

IV. CONCLUSIONS AND SUGGESTION

The results showed that: (1) Overall, there are differences in physics learning results between students are taught by GISA and students who are taught conventionally; (2) for a group of students who have a high learning interest, there are differences in learning results between taught by GISA and that taught by conventional; (3) for a group of students who have low learning interest, there are differences in learning results between taught by GISA and that taught by conventional; (4) There is no interaction between the learning model and learning interest toward the physics learning results of students.

Based on the research results obtained, it is advisable that matters: (1) For the teacher to more carefully select innovative learning model; (2) for students to be more daring demonstrate the ability and develop ideas held without fearing one inside the student to obtain optimal results; (3) For researchers who are interested in further elicits this study, is expected to examine the limitations in this study, so further research can enhance this research.

REFERENCES

- [1] Andrade, H. & Y. Du, Student responses to criteriareferenced self-assessment. *Assessment & Evaluation in Higher Education*, 2007, pp:159–181.
- [2] Basnet, B., Basson, M., Devine, J., Hobohm, C. & Cochrane, Is self-assessment effective in enhancing student learning? *Proceedings of the 2011 AAEE Conference*, Fremantle: Western Australia, 2006.
- [3] Gie, How to Learn Efficient [Cara Belajar yang Efisien]. Liberti: Yogyakarta, 1998.

- [4] Mansyur, Cultivating Responsibility, Confidence, Motivation and Mechanical Engineering Student Awareness Through Self Assessment [Menumbuhkan Tanggung Jawab, Percaya Diri, Motivasi dan Kesadaran Mahasiswa Teknik Mesin Melalui Self Assessment]. Jurnal PTM, 2005, pp: 57-64.
- [5] Nbina, J. B. & Viko, B.. Effect of Instruction in Metacognitive Self-Assessment Strategy on Chemistry Students Self-Efficacy and Achievement. Academia Arena, 2010, pp:1-10.
- [6] O'Malley, J. M. & L. V. Pierce, Authentic Assessment for Engglish Language Learners. Addison-Wesley Publishing Company: United State of America, 1996.
- [7] Ross, J.A. The Reliability, Validity, and Utility of Self-Assessment [Practical Assessment Research & Evaluation <http://pareonline.net/pdf/v11n10.pdf>], 2006, pp:10.
- [8] Rusman, Models of Learning Developing Teacher Professionalism [Model-Model Pembelajaran Mengembangkan Profesionalisme Guru]. PT. Rajagrafindo Persada: Jakarta, 2010.
- [9] Slameto, Learning and Factors Affecting [Belajar dan Faktor-faktor yang Mempengaruhi]. Rineka Cipta: Jakarta, 2003.
- [10] Sund, R. B. & L. W. Trowbridge, Teaching Science by Inquiry In The Scondary School. Bell & Howell Company: Ohio, 1973.

EFFECT OF COGNITIVE AND EMOTIF TECHNIQUES IN COUNSELING RATIONAL EMOTIF BEHAVIOR THERAPY TOWARD TENDENCY AGGRESSIVE BEHAVIOR BASED ON TYPE OF PERSONALITY AMONG STUDENTS OF SMP NEGERI 4 DENPASAR

I Wayan Susanta

IKIP PGRI BALI, Street Seroja East Denpasar Bali
iwayansusantadr@gmail.com

ABSTRACT- This study aims to determine the effect of the interaction between the techniques of cognitive and emotive techniques in counseling rational emotive behavior therapy against the trend of aggressive behavior in terms of the type of personality of students. The research was conducted in SMP Negeri 4 Denpasar. The study design used treatments by level 2X2. The independent variables in this study is the technique of cognitive and emotive technique. Cognitive techniques given to the experimental group and emotive technique given to the control group. Personality type plays a role as moderator variables which are divided into Personality introvert and extrovert. The dependent variable is the tendency of aggressive behavior. The research instrument consists of instruments personality types to distinguish students with introvert and extrovert personality, and instruments to measure the tendency of aggressive behavior tendencies aggressive level students. These samples included 108 students of class 8 by using purposive sampling technique sampling. Data analysis using analysis of variance of two lanes and Scheffe's t-test. The results show 1) there is a difference between the client's aggressive behavioral tendencies that follow cognitive techniques with clients who follow emotive techniques. 2) There is an interaction effect between counseling techniques therapy rational emotive behavior and personality type to the tendency of aggressive behavior. 3) cognitive technique is more effective in reducing the tendency of aggressive behavior on the client with extroverted personality, 4) emotive techniques more effective in reducing the tendency of aggressive behavior on the client with an introverted personality. Therefore it can be concluded that cognitive and emotive technique and personality type significantly influence declining trend Aggressive Behavior among Students SMP Negeri 4 Denpasar.

Keywords: *cognitive and emotive technique, personality types, and aggressive behavior*

I. INTRODUCTION

Secondary school students in their development require different understanding of themselves in relation to adaptation and social interaction. Self-understanding will always give you an idea about the weaknesses and advantages, as well as the acquisition of the ways of effective and efficient in the framework of self-actualization. The process of self-actualization will run very easy for them to adapt and it will be very difficult for those who are hard to adjust (mal-adjusted). The impact of a negative adjustment will cause various imbalances that can lead to acts of violence such as the emergence brutality or aggressive behavior tendencies.

Aggression can typically be defined as any form of behavior intended to hurt or harm someone who is opposed to the willingness of that person. Aggression involves any form of psychological or emotional abuse (Krahe, 2001: 5). Aggressive behavior by naturalists are all around us, not only in reality, but there is in the entertainment that we enjoy throughout the day-to-day. Impressions body part injured, news mutilation and a variety of sports that broadcast cumulatively violent elements will form the seeds of aggression among the television audience, especially among adolescents.

Judging from the result of aggressive behavior can cause physical injury, psychological, social, personal integrity, as well as damage to the object or the social environment. Further described aggressive behavior includes not only the physical aspects of the action, but also includes verbal abuse aimed at hurting others, psychological and symbolic or a combination of all these aspects. Verbal abuse such as insults and threats, psychological violence as attitudes that deny equal rights and human and symbolic violence that actions that cause fear and hostility.

Aggressive behavior itself is problematic behaviors that have an irrational belief. This behavior can be explained by the methods ABC Albert Ellis. A is antecedent event that all outdoor events experienced by the individual. Events of the past in the form of facts, events, behavior or attitudes of others. All of this is a trigger subsequent behavior. B is the behavior and belief. Behaviors and beliefs. There are behaviors based on the rational belief, there is also a behavior based on the irrational belief. Rational belief is a way of thinking or belief systems appropriate, reasonable and prudent. While the irrational belief is a way of thinking or belief system is wrong, irrational and emotional. Aggressive behavior is irrational behavior. While C is emotional consequence which is the emotional consequences as a result of the reaction of individuals in shaping a sense of excitement or emotional barriers in conjunction with antecedent event (A). The emotional ramifications of this is due to the accumulation between A and B / irrational Belief (Soekadji, 1983).

Scheneiders further states that aggressive behavior is emotion in reaction to the failure of an individual that is displayed in the form of vandalism against people or objects with the element of intent expressed in words (verbal) and non-verbal behavior. Aggressive by Murry defined as a way to fight with a very strong, fighting, wounding, attack, kill, or punishing others. Or in short aggression is an act intended to harm others or damage the property of others. aggression is not only done to harm the victim physically, but also a psychic (psychological), for example through the activities of insulting or blame. (Nadhirin, accessed 1 February 2012).

Aggressive behavior among teenagers appear to have happened in a fight between teenagers, bullying measures, and various acts of violence involving youth, such as violent motorcycle gangs in various areas such as; in Medan, Jakarta, Bandung, and other areas. In Denpasar-Bali Denpasar Seseetan particularly in the area south of persecution on young girls by a Gang of young women (Mantra, "Bali Post, February 11, 2012: 2). The emergence of aggressive behavior based on research results from Fajar Purnomo Askuan, showed that the factors of the background for the students to behave aggressively due to economic circumstances, parenting, modeling, personal, social causes and the social environment (Askuan, accessed 28 January 2013) ,

"The most fundamental causes of aggressive behavior like this is adverse environmental effects, such as relationships with friends (peer group), a lack of discipline cultivation, or the wrong upbringing of parents, for example educating children too hard or too soft. (Asmani, 2012: 36). This is in line with the theory of aggression in an environment that is social learning theory. Social learning theory argues that aggressive behavior is a learned behavior by watching others acting aggressively. Basically, human beings are not born with a number of aggressive responses but they must obtain this response by way of experiencing it directly (direct experience) or by observing other people's behavior. The triggering of emotions because of the unpleasant experiences can motivate aggressive behavior. In predicting the likelihood of behavior that has been studied to actually be implemented, perception due to the behavior of the model and the observed behavior and painful experience plays an important role. Behavioral models and consequently an external stimulus that elicits aggressive response tendencies on the part of the observer. Normative standard observer, as well as the adequacy of the observed behavior and their self-efficacy (ie the belief that he was able to show the behavior with the desired effect), the internal mechanisms that regulate aggressive behavior. Social learning perspective has made important theoretical approach to conceptualize the effects of media violence to aggressive behavior, which can be considered as a paradigmatic case of learning through observation and reinforcement from others (Myer, 2011: 70).

Aggressiveness is a motif that is present in the life of every individual, as well as the intensity, quality and manifestations may differ between one individual with other individuals. Facts show fight occur between students in the area Seseetan and child stabbing junior high school children (Bali Post, 11 February 2012). Factual matters which is an indication of the tendency of aggressive behavior among teenagers in general.

Seeing this as above, as education practitioners we must not stand idly by. We will undertake prevention measures through research activities to sensitize teenagers quasi from view irrational view of the underlying attitudes and behavior. Counseling for the Rational Emotive terafi behavior will be the approach to change the views of the irrational becomes rational adolescents with cognitive techniques and emotifnya.

Rational Emotive terafi behavior counseling focused on solving problems that focus on aspects of thinking, judging, decided, without a more directive dealing with the dimensions of the mind rather than the dimensions of feeling (corey, 1988). Rational Emotive terafi behavior counseling can be seen as a model of cognitive behavioral therapy oriented. This approach is oriented in such a way, which eventually transformed into a comprehensive and eclectic approach that emphasizes the elements to think, reason, decide and do. Rational Emotive Behavior Therapy Definition belong to the cognitive-oriented counseling. This approach is one form of active-directive counseling that resembles the teaching process, persuasive, confrontation and administration tasks, while maintaining the dimensions of the mind rather than feelings. Trying to eliminate the way of thinking clients that are not logical and rational becomes irrational, eliminate emotional disorders, build self-interest, self-direction, tolerance, acceptance of uncertainty, flexible, commitment, scientific thinking, risk taking, and self-acceptance of clients. (Surya, 2003).

The application of these techniques in counseling rational emotive behavior therapy (REBT) can be adapted to the type of personality the client to achieve the effectiveness of the treatment. Type personality distinguished clients become introverted and extroverted personality. "Gregory G Young" explained the characteristics of introverted personality is quiet, shy, introspective, love reading, aloof and distant except with a friend who is already familiar, tend to plan ahead - to see first - before moving, and suspicious, Like the excitement, live a daily life with seriousness, and love the lifestyle are well organized, keeping his feelings behind closed doors, rarely behave aggressively, do not eliminate anger, trustworthy, in some ways pessimistic, and has a value of high ethical standards. While extroverted personality is one end of the personality dimensions of introversion-extroversion with characteristic gentle disposition, outgoing, friendly, love impulsive, and likes to take risks. The characteristics of extroverted personality is sociable, likes parties, has many friends, needs a friend to talk to, and do not like reading or studying alone, craves excitement, take up challenges, often against danger, act without thinking first, and usually impulsive, fond of a joke-joke, always ready to answer, and usually like to be a change, carefree, not much consideration (easy going), optimistic, and love to laugh and happy, prefers to remain engaged in activity, tend to be aggressive and quickly lost his anger, all his feelings are not kept under control, and not always reliable (Young, 2008: 19).

On the basis of techniques studies in counseling and personality type on counseling techniques can be adapted to the type of personality of students. Students with more appropriately counseled introverted personality with emotive technique because it is covered more precisely with sociodramas techniques, self modeling and assertive training in expressing feelings. While students with more appropriately counseled extrovert personality with cognitive techniques due to the open nature of the teaching techniques, persuasive, confrontation and giving the task to understand the things that are irrational.

Based on the above description can be formulated several research problems as follows. (1) Is there a difference between the aggressive behavior tendencies of students who counseled with cognitive techniques with students who counseled with emotive techniques. (2). Is there an interaction effect between cognitive techniques and techniques emotive and personality type on the trend of aggressive behavior (3) Are there differences in the tendency of aggressive behavior among students who counseled with cognitive techniques and students are counseled by techniques emotive, the students who have personality types extrovert (4) Is there a difference between the aggressive behavior tendencies of students who counseled with cognitive techniques and students who counseled with emotive techniques, the students who have introverted personality types.

While the purpose of this study were (1) to determine differences in the tendency of aggressive behavior among students who counseled with cognitive techniques with students who counseled with emotive techniques. (2). to determine the effect of interaction between cognitive techniques and techniques emotive and personality type on the trend of aggressive behavior (3) to determine differences in the tendency of aggressive behavior among students who counseled with cognitive techniques and students are counseled by techniques emotive, the students who

have personality types extrovert (4) to determine differences in the tendency of aggressive behavior among students who counseled with cognitive techniques and students are counseled by emotive techniques, the students who have introverted personality types.

The results of this study are expected to contribute positively to the development of educational science, psychology, counseling, educational evaluation and research primarily as a reference for the conduct of relevant research, in addition to the benefits of the theoretical, the results of this study are expected to provide practical benefits for the managers of education , By knowing the tendencies of aggressive behavior of students and know ways to prevent it through the techniques of proper counseling, would be useful for students, teachers BK, parents and decision makers in order to bring other approaches preventive - educative as a solution to the phenomenon aggressiveness.

II. RESEARCH METHOD

This study used a quasi-experimental design with treatment by level design. Level A includes REBT counseling that includes techniques Cognitive and Emotive techniques. While the level of B include the type of personality of students that includes extrovert and introverted personality. The research was carried out among students of SMP Negeri 4 Denpasar Bali province. The study was conducted in January 2016. The sample of 108 people taken by purposive sampling technique. Data aggressive behavioral tendencies and personality types of students were collected by questionnaire previously tested validity and reliability. Data analysis technique used is the analysis of variance (ANOVA). The hypothesis tested was hypothesized main effect (main effect), the hypothesis of interaction (interaction effect), and hypotheses (simple effect).

III. RESEARCH FINDINGS AND DISCUSSIONS

Results of ANOVA two-lane calculation shows that the value of F between the level of factor in the type of counseling techniques (between columns) F count obtained was 9.99 while price F table for DKA = 1 and DKD = 104 at the 0.05 significance level of 5.175. Turns F count larger than F table (of $F = 9.99$ $F(0,025)(1: 104) = 5.175$). This means that the null hypothesis (H_0) that states that there are differences in the tendency of aggressive behavior among clients who were counseled by emotive techniques and counseled clients with cognitive techniques, rejected. In contrast, the alternative hypothesis (H_a) states that there are differences in the tendency of aggressive behavior among clients who were counseled by emotive techniques and counseled clients with cognitive techniques, accepted.

Besides, it was also found interaction between rational emotive behavioral counseling techniques therapy (emotive and cognitive techniques) and the type of personality to the tendency of aggressive behavior of clients. Tests conducted on statistical hypothesis is formulated as follows.

$$H_a: \text{INT. } A \times B \neq 0$$

$$H_0: \text{INT. } A \times B = 0$$

The results of analysis of $F = 41.03$, while the price for the F table and the 5% significance level: $2 = 0.025$ by 5.175. This means, that H_0 rejected and H_a accepted. It can be concluded that there is interaction between the type of rational emotive behavioral counseling techniques therapy (emotive and cognitive techniques) with the type of personality to the reduction of aggressive behavior tendencies clients.

The results of data analysis showed that a group of clients who have extroverted personalities who counseled with emotive technique (A_1B_1) had a propensity score of aggressive behavior by an average of 80.15, while a group of clients who have extroverted personality and cognitive techniques are counseled to follow the cognitive techniques (A_2B_1) has a propensity score of aggressive behavior by an average of 64.85. Meanwhile, the results of the calculation of two-lane ANOVA showed that average Squares In (RJKd) of 69.01. The next test t-Scheffe for determine which group is superior. From t-Scheffe test results obtained at 6.8 while table to 0.05 significance level of 2.00.

From t-Scheffe test showed that more than table count. This means that the null hypothesis (H_0) stating that for clients who have extroverted personality who was counseled by his aggressive behavior tendencies emotive technique is not better than those counseled clients

with cognitive techniques, rejected. In contrast, the alternative hypothesis (H_a) states that for clients who have extroverted personality, aggressive behavioral tendencies clients who were counseled by emotive technique lower than those counseled clients with cognitive techniques, is received (fail rejected). So it can be concluded that the clients who have extroverted personality, and cognitive techniques aggressive behavior tendency lower than those counseled clients with emotive techniques.

The results of data analysis showed that a group of clients who have an introverted personality and counseled with emotive techniques (A1B2) had a propensity score of aggressive behavior by an average of 66.63, while a group of clients who have counseled with extrovert personality and cognitive techniques (A2B2) has a propensity score aggressive behavior by an average of 71.82. Meanwhile, the results of the calculation of two-lane ANOVA showed that average Squares In (RJKD) of 69.01. Furthermore, the Scheffe's t-test to determine which group is superior. From t-Scheffe test results obtained at 2.29, while for a significance level of 0.05 table 2,000.

From t-Scheffe test showed that t count greater than t table. This means that the null hypothesis (H_0) that states that the client has an introverted personality, aggressive behavioral tendencies lower for clients who were counseled to follow a cognitive technique is not better than the tendency of aggressive behavior of clients who were counseled emotive, rejected. In contrast, the alternative hypothesis (H_a) which states that the clients who have an introverted personality, aggressive behavioral tendencies clients who were counseled by emotive techniques lower the tendency of aggressive behavior were counseled clients with cognitive techniques, accepted. So we can conclude that for clients who have an introverted personality, aggressive behavioral tendencies clients who were counseled by emotive technique is lower than the tendency of aggressive behavior were counseled clients with cognitive techniques.

Based on the above research findings, there are several things that can be delivered, the first that the clients with extroverted personality has the openness to social dynamics. For the change counseling therapy rational emotive behavior seems to make a considerable contribution both as reported in the Journal of abnormal psychology Vol 79 (1) Feb, 1972, 60-67. <http://dx.doi.org/10.1037/h0032336>, by Trexler, Larry D.; Karst, Thomas O., stated that REBT is more effective than no treatment or relaxation to be used to eliminate anxiety / emotional consequence due to the indirect influence of various factors antecedent in addition to the influences of self belief irrational clients. (<http://psycnet.apa.org/journals/abn/79/1/60/>, accessed May 25, 2016).

Similarly, in the journal behavior therapy Volume 8, Issue 4, September 1977, Pages 567-574 with a report with the title: Modifying assertive behavior in women: A comparison of three approaches, modification of assertive behavior in women, a comparison between the three types of approaches therapy among other things, (1) modeling -plus - exercise behavior (behavior rehearsal), (2) modeling -plus - -plus exercise behavior - rational therapy (RBT), a consciousness-raising groups (CR) a consciousness-raising group, and (WL) non treatment a waiting list control than in the treatment of difficult to express an opinion, the treatment given to 64 women in outpatient clinical settings. Second BR (Behavior Rehearsal) and RBT (rational behavior therapy) showed a significant improvement in the content and scale of the size paralinguistic behavior in situations treated. This effect can also be generalized to the situation treated. Only RBT showed a decrease in situational anxiety. While CR / consciousness-raising proved ineffective.

(<http://www.sciencedirect.com/science/article/pii/S0005789477801871>, accessed May 25, 2016).

In the introverted personality, therapy emotive technique is highly effective in reducing anxiety as reported in the Journal of the Indian Academy of Applied Psychology, Volume 8 Number 1, Pages 169-173, through a quasi-experimental study that counseling rational emotive behavior terafi effective to remove the shame specialized in among adolescents. Teenager boys is higher than the reduction in their shame young women. ([Www.jiaap.org/listing.../ba65b0f5-6158-4c3d-b135-28bed1573ced.pdf](http://www.jiaap.org/listing.../ba65b0f5-6158-4c3d-b135-28bed1573ced.pdf), accessed May 25, 2016).

In addition to the above, Razieh Kamal also examined the influence of counseling rational emotive therapy to self-esteem and aggressive behavior as reported in the journal Psychology volume 5, issue 10 October 2015, with the title Research impact of rational emotive therapy on self esteem and aggression Among adolescent students , The findings of the study

states that rational emotive counseling therapy effective to increase self-esteem and lower levels of aggressiveness among teenagers India.

([www.worldwidejournals.com/ijar/file.php?val=October_2015 ... 47 ...](http://www.worldwidejournals.com/ijar/file.php?val=October_2015...47...) accessed May 25, 2016).

The findings of the study above shows the effectiveness and consistency of rational emotive behavior therapy counseling in tackling the problem among teenagers. So that kind of terafi like this has a value of decency for the officers involved to try to implement it.

IV. CONCLUSIONS AND SUGESTIONS

The findings of this study is for clients who have a more extroverted personality type fits counseled with cognitive techniques, while for clients who have introverted personality types more suited counseled with emotive techniques. In order to improve the quality of the learning process and minimize the tendency of aggressive behavior of clients, suggested to the guidance and counseling teachers, counselors and practitioners use techniques related to cognitive and emotive appropriately and correctly.

REFERENCES

- [1] Agus Daryo. 2009. Parenting Parents. Jakarta: Ghalia Indonesia
- [2] Asmani Jamal Ma'mur. 2012. Tips Delinquency in Schools. Yogyakarta: Blue Book.
- [3] Dantes, Nyoman. 2014. Analysis and Design Eksperimen. Singaraja: Graduate Program Undiksha.
- [4] Daryoush Ghasemian, Lancy D'Souza and Sedigheh Ebrahimi, Journal of Indian Academy of Applied Psychology, Volume 8 Number 1, Pages 169-173, India: University of Mysore. (www.jiaap.org/listing.../ba65b0f5-6158-4c3d-b135-28bed1573ced.pdf, accessed May 25, 2016).
- [5] Gerald Corey. 1988. theory and practice of counseling. Bandung : Gramedia
- [6] Janet L. Wolpe, Iris G. Podor. Modifying assertive behavior in women: A comparison of three approaches, Thrapy Behavior Journal, Volume 8, Issue 4, September 1977, Pages 567-574. <http://www.sciencedirect.com/science/article/pii/S0005789477801871> (accessed May 25, 2016).
- [7] Krahe, Barbara. The Social Psychology of Aggression, translation Helly Prajitno Soetjipto. Yogyakarta: Pustaka Pelajar, 2001.
- [8] Mantra Dharmawijaya. "Concerned Youth Against Violence." Bali Post, February 11, 2012.
- [9] Nadhirin. Behavior-Aggressive-Teen, <http://nadhirin.blogspot.com/2009/12/.html> (accessed 1 February 2012).
- [10] Rasieh Kamal, Venkatesh Kumar, G. psychology journal volume 5, issue 10 October 2015, *the impact of rational emotive terapy on self essteem and aggression Among adollescent students*. India: University of Mysore ([www.worldwidejournals.com/ijar/file.php?val=October_2015 ... 47 ...](http://www.worldwidejournals.com/ijar/file.php?val=October_2015...47...) accessed May 25, 2016).
- [11] Soekadji Soetarlinah. 1983. *behavior modification, application of daily and professional application*. Yogyakarta : Liberty
- [12] Surya, Mohamad. *basic elementary education counseling, concepts and theories*. Bandung.
- [13] Trexler, Larry D .; Karst, Thomas O. Journal of Abnormal Psychology, Vol 79 (1) Feb, 1972, 60-67. <http://dx.doi.org/10.1037/h0032336> (accessed May 25, 2016).

THE EVALUATION OF THE SCHOLARSHIP DEGREE PROGRAM FOR THE ISLAMIC RELIGIOUS EDUCATIONAL TEACHERS AT SCHOOL

Ju'subaidi

Lecturer of Ponorogo Islamic State Collage

Email; *subaidi_6@yahoo.com*

Abstract - The aim of this research to evaluate of implementation of Qualification scholarship degree for The Islamic Religious Education Teachers at the school. This research is performed in Ponorogo Islamic State College and used the qualitative approach. The data is gathered with interview and documentation methods. It used the interactive to analysis data consist of data reduction, data display, and conclusion. The result of the research are; preparation of implementation have been only performed physically. Implementationn program is not according to standart wich is decided by goverment. The side effect of the program implementation is misinterpretion to the program. Economically, inevisien of bugeting, and Instution does not have quite useful enough from the program.

Keywords: *Qualification Scholarship degree, Islamic religious education, misinterpretation, inevisien.*

I. INTRODUCTION

Improving the quality of education in autonomy is determined by various factors, such as educational resource factors and educational resources management education. Educational resources regard to funding, the quality of teachers, and instructional media.

Teachers as education component plays a key role in improving the quality of education and are in the central point of any education reform efforts. There are not a few a teachers who do not have competence in performing their duties. Therefore, teachers need to be developed their qualifications and competence. Judging from his qualifications, there are still many who do not meet the conditions set by law teachers, especially at the level of both the Primary School and the Islamic Elementary School (MI).

Based on data from the development of the number of teachers in 2008 (submanpower/Ditpa), that the number of teachers of Islamic education in schools from kindergarten to high school teachers amounted to 168.184. From these teachers there were 1,539 Strata-2/-3, 80,134 S-1 /D-IV, and 86.511 people have not qualified teachers D-IV/S-1. In other words, 51,438% of total Islamic teachers who have not qualified S-1.

One of the efforts to realize the mandate of law No. 14 of 2005 on teachers and lecturers and Government Regulation No. 19 Year of 2005 on National standards Educational updated by Government Regulation No. 32 Year 2013 is to perform the education profession that allows the teachers to master intact competencies, which is expected to contribute to the improvement of the quality of education.

Noting this, it is required an innovative and efficient initiatives to provide educational services that allow not disruptive to the daily tasks of each teacher. Correspondingly, the Directorate General of Islamic Education Ministry of Religious Affairs of the Republic of Indonesia started the academic year 2009/2010 has issued policies to administer the scholarship program Strata (S-1) for Islamic Education teachers serving in schools.

Ponorogo Islamic State College as one of the Islamic religious higher education in the Ministry of Religion of the Republic of Indonesia has been appointed to implement the scholarship program Undergraduate 1 for teachers in school who do not have or qualify for S-1.

As an institution for the program, Ponorogo Islamic State College is required to manage the program with full responsibility. To find out how high the quality or condition of something as a result of the implementation of the program and the level of achievement of the objectives of the program and the things that support or inhibit, the following program will discuss implementation of the program.

II. METHOD

This study is an evaluative research with a qualitative approach. This research is categorized into qualitative research because they are around the globe, researchers as a key instrument, the world trying to uncover the meaning behind a person's actions (Bogdan & Biklen, 1998: p.4-7). Evaluative study is intended to look at the effectiveness of the implementation of a program (Joint, 1981: p.12). While the program in question is a scholarship program S-1 Degree qualification for teachers of Islamic religious education in schools to improve the qualifications and competence. This study tried to look at the process of implementation of the above program and see the positive and negative effects posed by the program.

Evaluation model used is the model evaluation Goal Free Evaluation that introduced by Michael Scriven. This model emphasizes the impact of the implementation of the program of the objectives of the program, as stated by Michael Scriven, "Goal-Free Evaluation after noticing side effects that sometimes had a more positive (or negative) effect than did the intended goals". This theory identifies the impact of the implementation of the program.

This research was conducted in Ponorogo Islamic State College as program implementers did not touch on policy director General of Islamic Education Ministry of Religious Affairs as the authority program. It is intended that the organizers are assumed to have got a guide implementation. The subject of research is the program and program participants amounted to 18 people. Data were collected through interviews and the structure is not supported by documentation technique, then analyzed by interactive analysis includes reduction, display, and conclusion (Sugiyono, 2006: p.267). Validation of the accuracy of the findings with peers, triangulation of sources, and the extension of the observation.

III. THEORICAL

There are three words that are often used overlapping: the measurement, evaluation, and assessment. Measurement is defined as a process of determining the number of individuals and or characteristics according to defined rules (Denzin, 2000: p.769). While Allen and Yen in Djemari Mardapi, defines the measurement as a process of determining the number of individuals and or characteristics according to defined rules (Mardapi, 2008: p.1). While the assessment is the process of providing information on all matters relating to the institution. Stark and Thomas define the assessment is "processes that provide information about individual students, about curricula or programs, about institutions, or about entire system of institutions" (Stark & Thomas, 1994: p.46). While the definition of evaluation as a systematic action of social research procedures to measure the conceptualization, design, implementation, and advantage a program (Rossi & Freeman, 1993: p.5). Cronbach stated that keyword evaluation process is a little proof of the effectiveness of the use of funds and the program runs according to plan (Cronbach, 1982: p.1). Worthen explained that a good evaluation always utilizes measurements and observations are accurate, reliable, and valid for systematically collecting empirical evidence is then used to analyze the results objectively (Worthen & Sanders, 1987: p.98).

The main objective of the evaluation study by Ogle, is to find the specific information needed to make a decision, not just to obtain information about effectiveness program (Ogle, 1982: p.3). Evaluation research is a method to learn something that is complex from the implementation of a program based on a comprehensive understanding of the description and analysis (McLaughlin & Jordan, 1999: P.72).

In connection with the purpose of evaluation research that is intended to improve the program, justification whether a program was successful or not in achieving the desired objectives, the necessary standards (benchmarks) that can be used as a reference of justifications. Therefore, the standard used in this study is the existing guidelines in the handbook academic Ponorogo Islamic State College in 2012.

Valades & Bamberger stated that evaluation is internal or external management activity that is intended to measure the feasibility of a program design and implementation methods (Valadez & Michael, 2007: p.283-296). Evaluation of the program as a systematic method to measure and

improve a program through a procedure that is ethical, accurate and feasible (Sanders & Fitpatrick, 1996: p.24). McLaughlin & Jordan, evaluation program is a systematic process of gathering information about the activities, characteristics and outcomes of the implementation of a program with the intent to make repairs and determine the effectiveness of the program and make decisions on the sustainability of the program (McLaughlin & Jordan, 1999: p.65).

While the Joint Committee on Standards for Educational Evaluation states "program evaluations that assess educational activities which provide service on a continuing basis and often involve; curricular offerings". Valadez and Bamberger assert that the information obtained through the evaluation activities of this program can be used to measure the implementation and impact of the program.

Starting from some of the above understanding, program evaluation can be understood as the activity of systematically gathering information on the implementation of a program intended to determine the feasibility of the program in order to achieve the desired objectives and impact.

Evaluation of Non-purpose (Goal Free Evaluation) has been introduced by Michael Scriven stated, "Goal-Free Evaluation after noticing side effects that sometimes had a more positive (or negative) effect than did the intended goals" (Ogle, 1982: p.29). This theory is based on the emergence of the impact of the implementation of the evaluation that often arises out of the program objectives. The evaluation needs to be considered is how the program works, by identifying the performances that happened, good things are positive (expected) and negative things (that are not expected (Arikunto, 2008: p.41).

Therefore, in implementing the evaluation program, an evaluator does not need to consider what the objectives of the program, because sometimes the side effects are even more positive value than the intended purpose. If each of the specific objectives achieved, meaning that met in appearance, but the evaluators forget to consider how far each of these performances are expected to support the final appearance by a common goal then consequently the amount of this special appearance will not be much benefit.

A goal-free Evaluation focuses on the actual results of a program or activity. As a result, goal-free evaluation increases the likelihood of undesirable results will be identified and recorded (Subarsono 2009: P.20). The use of goal-free model evaluation, the evaluators need to produce two items of information, namely; a) assessments of the real (actual effects), and b) an assessment of the profile needs to be assessed (Sukardi, 2000: p.61).

In other words, if a product has an influence that can be demonstrated in a real and responsive to a need, means a planned product useful and positive way need to be developed. Interpretation opposite occurs, if a product, including teaching and learning activities, have no real influence on the participants then it does not need to be developed.

A. The Scholarship Degree Program for Islamic Education Teachers at School.

The Scholarship Degree Program for Islamic Education Teachers at the school in the form of lectures Special Regular of Minister of Religion fiscal year 2010. The program is the provision of funds to GPAI on schools to follow the S-1 education at the College of Islamic Religion of State (STAIN) Ponorogo (MoU, article 2 (1): p.3).

This program is an effort to improve the quality of Islamic education in schools and supported by the availability of human resource educators who have an undergraduate qualification with competence Islamic religious education, thereby producing learners who qualified, both aspects of the faith, piety and noble character. Matters relating to the implementation of this program managers use a benchmark in the book's academic guidelines Ponorogo Islamic State College in 2010.

Graduates Competency Standards which set is having pedagogical competence, personal competence, professional competence and social competence, and leadership competence. While academic qualification is Strata (S-1) in Islamic Education.

The structure of the curriculum program was form in a special regular lectures at Ponorogo Islamic State College. The curriculum follows the course of Islamic Religious Education (PAI) Ponorogo Islamic State College, with the number of credits that must be taken by participants of this program. It is 144 credits with the study of 8 (eight) semesters for which high school education (Podoman Academic, 2012: p.13). While a Diploma two (D II) is 104, and the course duration of six semesters. Learning was the system includes a model of learning and

assessment. Learning is done on this program uses a special regular approach so that participants do not leave their duties teaching at each school

Indicators of program success is declared if frequency-face to reach 75% and scored 75 consistently. As for monitoring the implementation of the programs carried out by the monitoring and evaluation team composed of representatives from the Directorate General of Islamic Education and the Directorate of Islamic Education in Schools. This providing program is monitored by the Chairman of Ponorogo Islamic State College as well as a charge of the program.

IV. FINDINGS AND DISCUSSION

A. Preparation

Legally, the scholarship degree program have gained legality to be implemented through the Minister of Religious Affairs of Indonesia number 179 Year 2008 on the implementation of Improvement Qualification Bachelor (S-1) for Teachers PAIS. State Islamic Institute (STAIN) Ponorogo is one of the State Islamic University (PTAIN) which is trusted to hold the program. Geographically, Ponorogo Islamic State College actually has a large area covering five districts and one city Associate. This program was attended by 67 participants from four districts namely Ponorogo, Pacitan, Magetan, Madiun and Madiun.

The authority of the implementation of this programore will be more completed when managers were given full authority since the registration of participants. Based on the findings, the executor is only authorized to give information to the public and the administration selection. The involvement in the selection of candidates at the pre-implementation program will provide objectivity in the selection of participants so that the participants will minimize data manipulation. Program participants were designated as permanent participants in Ponorogo Islamic State College totaling 67 people. There are some participants who actually do not meet the requirements to join the program. A previous participant was not a remain teacher from a school. Because these participants have a close relationship with the board of trustees of the Institution who issued the Certificate, the certificate of institution can be obtained easily. Therefore, a prerequisite set by the center office in this regard Ditjenpais equipped with tenure candidates at least 5 years, it will eliminate the impression of taking over the access of others to gain increased academic qualifications and are more entitled to accept the scholarship.

As with the two participants who had more age limit of 48 years at the time of registration as listed in the guidelines of Academic Program Implementation Scholarship Undergraduate (S-1) GPAI (Islamic religious educational teachers) at school in Forms of Regular special Class in Ponorogo Islamic State College article 9, paragraph 3 (b), the manager and the Directorate of Islamic Education has violated the rules he made.

Of the two participants, one person was 49 years old in 2010, and the other was 51 years old in 2011. This situation is compounded by the status of the participants from the two above-mentioned participants as Principal and has had academic qualifications Strata One (S-1) in Bahasa Indonesian. Participants are thus more likely caused by the process of recruitment of candidates who are less careful and rules that are used as the legal law provide a loophole for engineered and manipulated. On the other hand, managers rely more on physical documents are not validated as an effort to ensure the accuracy of documents submitted. If the participant is surveyed to a job candidate in advance, it would provide data support that objective and reduce fraud and manipulation. When the device is used as a reference law contains clear and detailed limits would reduce risky deeds which contrary to legal norms and social norms. In effect, the legal instruments and clear rules and detailed will not necessarily be able to guarantee both in terms of implementation of the program. However, with the clarity of the rules will help managers facilitate the selection, as the manager has always argued that the authorities of the release candidate is the center office. If so, it is necessary to support the attitude of commitment (political will) the managers to use rules or norms of law in running the program. Another impact is occurring on the part of government and school/madrasah conferred a certificate harmed, if the concered candidates do not teach proposer instituted after the completion of the program

Participants of the program from kindergarten teachers will also be part of the factors inhibiting the achievement of the objectives of the program, considering one of the functions is the usefulness of the program. Hire teachers and park-school believed to be less having value as an institution, because the courses taken by participants from kindergarten to the material to be

provided to participants in kindergarten irrelevant, although personally for participants to have a value added.

Observing the subjects listed in the guidebook academic article 16, the curriculum used by the manager is a curriculum for Islamic religious education courses primary and secondary levels. Subjects in the curriculum does not show on the territory of the Early Childhood education or kindergarten. In terms of curriculum between courses of Islamic religious education in kindergarten with elementary and secondary have a considerable difference. The difference is caused by differences in the level of thinking and development psychology among learners at pre-school to primary and secondary levels. In this case, should the material and lectures among teachers from kindergarten to the elementary and secondary levels are separated.

Noting courses were given to the teachers in the program are not correspond to their expertise. Non-compliance will lead to the delivery of material is not believed to be the maximum, because they are less scientifically controlled. The condition was aggravated by lifting Lecturer (DLB), which has not had a teaching certificate as required by law teachers and lecturers, one of whom got an administration that legally do not have the authority as a lecturer. Perspective of Law No. 14 year 2005 that the evidence as a lecturer is their ownership Certificate is educator will explain areas of expertise. The government's efforts certainly not just an educator's certificate, but more towards the job descriptions of the academically indicated on the professionalism of a lecturer. This seems less a concern for program managers. The manager assumes that every lecturer is able to support any subjects. If this becomes a rationale for the provision of courses taught by lecturers who are not his field, the participants believed the program will not earn a lot of material, because the professor also did not control many materials. If the manager would use a lecturer accordance with the field, learning problems can be reduced.

B. Implementation

In the perspective of management of the implementation program of activities for the realization of planning (planning) into tangible action in order to achieve program objectives effectively and efficiently. Planning has been prepared will have a value if implemented with appropriate and timely. Vice versa, if implemented with modest then what happens is has no meaning (useless).

In the implementation of any organization or program should have the power steady and convincing. Through good management program is expected to contribute directly to the achievement of the goals of running programs. Based on the data and findings of the implementation of the program in general is progressing according to the set time. The main objective of the program was completed academic qualification improvement until graduation as many as 42 participants. While the 23 participants were in the seventh semester, and still in the process of coaching techniques thesis

On the other side, the successful implementation of the scholarship program is in achieving its objectives, there are several issues that still require attention to be seriously improved. The issue is intended oversee the process of lectures conducted by outstanding faculty, because they have not proven competence. Recruitment of outstanding faculty impressed closeness but not in its competence. If the task of the manager is positioned as the main tasks and functions (TUPOKSI) managers should instead involve the lecturers still existing in accordance with the necessary competence.

Assessment as part of a series of lectures and become a measuring tool mastery of competencies students participating in the program should be a priority manager. Condition assessment exercise in the form of midterm and final semester of origin impressed implemented. Implementation assessment puts less valuation principles. The case of the occurrence of each other to provide answers to other friends when the midterm and end of term to indicate poor implementation of that judgment. Earned GPA which is on each card results of a study (KHS) has yet to show a score of acquisition and actual competence. In other words, the score seemed not depict the actual quality.

The financial management program has not shown good financial management. The indicator is the program manager at the beginning of obtaining honorarium, while the other half did not receive honorarium. By law managers should obtain honorarium as has been stated in the Decree of the Chairman of Ponorogo Islamic State College in the fourth clause. The impact is the manager taking unilateral policy to acquire honorarium by engaging themselves become lecturers of the subjects in the program.

It was 75% of Program managers who has position as structural officials. Aspects of the managerial capacities of the manager fails to demonstrate commitment to a low. It can be seen less program management in accordance with the guidelines of operations, since the recruitment of prospective participants of the program, faculty recruitment, implementation of lectures, curricula imposed and governance of finance. The condition is sociologically will affect the beliefs and sentiments of society to institute Ponorogo Islamic State College.

Based on the data obtained, scholarships disbursement between program managers from public and private universities have different treatment. Community college scholarship search lasted a year, while the private universities in the form block grant. The difference in treatment caused a reaction that is considered to be the cause of constraint the realization of a wide range of program support activities. For every time the managers will hold the supporting program, the managers must submit proposals as a condition for the disbursement of funds. As a result, many funds were not absorbed optimally and must be returned to the state. The disbursement models and procedures for using the funds paid to the treasurer Ponorogo Islamic State College actually depends on the manager of the program. That is a level of creativity to make the program support managers will affect the rate of absorption of the funds available, the more creative managers in making support activities, the more extensive program managers the opportunity to empower funds.

The Use of funds directly by students have been reported through letter Accountability should be made by each of the students participating in the program at the end of each academic year. There is a bit of a mistake in the use of indirect cost based on existing data. Financing of the founded Islamic school workshop activities guided by the management has violated procedures of use of funds, since funds should be allocated for teacher improvement program activity of Islamic religious education in schools transferred to the territory of the madrasah. If you pay attention to the clauses contained in the guidebook academic program delivery Scholarship Undergraduate S-1 GPAI at school article 9, paragraph 3 (a) on page 7 stated that the College organizers designated will register the candidate with the criteria, participants are teachers PAI in Kindergarten, Elementary, Junior high school or Vocational school

The above criteria shows that the scholarship program reserved for teachers of Islamic religious education who served in kindergarten, elementary, junior high, high school or vocational school is not in madrasah. If we look at hieararchy, Madrasahs are educational institutions that are not under the authority of the Directorate General of Islamic education at the school. Therefore the use of funds for madrasah built into misuse of funds, although to the two institutions are still under the Ministry of Religious institutions of the Republic of Indonesia.

C. Impact

Scholarship Qualifying Implementation GPAI at Ponorogo Islamic State College, since its implementation in 2010 up to the academic year 2013/2014 has resulted sociological impact, namely public who have high school educational background and Diploma II high school can continue to strata one. Economically, the direct distribution of funds are received by students participating in the program has helped to reduce spending as usual costs everyday. Most of the participants who have passed directly follow the certification, the certification has also gained benefits, but it should be noted that some of the funds should be for the benefit of the addition of a reference and a copy but it is used to buy a laptop. Another disadvantage is the manager did not get the honorarium of being perceived as the main tasks and functions. This will obviously detrimental to the managers that have an impact on the process of managing the program less than the maximum.

Judging from the academic impact, it is less build the capacity of participants in the art program. Besides there are injured parties, some parties who benefit from the implementation of the program. The advantage gained is certainly varied, from financial and non-financial advantages.

V. CONCLUSION

Preparation of implementation Scholarship Degree program for Islamic Education Teachers at the school, which made the new manager is limited to physical preparation. While the non-physical preparation in the form of understanding of the program's existence does not do. Implementation of the Scholarship degree program for Islamic Education Teachers at the school have a partially compliant program operations so that the program's goals is less than optimal. The impact was a miscommunication and misinterpretation on the performance of managers. On

the other hand, financial losses due to inefficiencies. Academically, the competence of program participants is not measured properly so the quality of program objectives are not achieved. On the other hand, institutionally less benefit from the program.

REFERENCES

- [1] Arikunto, A. (2008). *Evaluasi program pendidikan*, Jakarta: Bumi Aksara
- [2] Bogdan, R.C. & Biklen, S.K. (1998). *Qualitative research in education: An Introduction to theory and methods*. Boston: Allyn and Bacon
- [3] Cronbach, Lee J. (1982). *Designing Evaluation of Educational and social programs*, San Fransisco; Jossey-Bass Publishing
- [4] Denzin, K. Norman & Lincon., (2000). *Handbood of qualitative research 2nd*, Thousand Oaks, Sage Pablication
- [5] Joint Committee on Standards for Education Evaluation, (1981). *Standarts for evaluations of Educational programs, projects and materials*, New York, McGraw-Hill Book Company
- [6] Jurusan Tarbiyah STAIN Ponorogo., (2012). *Buku Pedoman Akademik*, Penyelenggaraan Program Beasiswa Kualifikasi Sarjana S-1 Guru Pendidikan Agama Islam pada Sekolah di STAIN Ponorogo, Ponorogo: Jurusan Tarbiyah
- [7] Mardapi, D. (2008). *Teknik pengyusunan instrumen tes dan nontes*, Yogyakarta: Mitra Cendikia
- [8] Mclaughlin, J.A., & Jordan, G.B., (1999). *Logic Models; a tool for telling your program's performance story*. *Evaluation and program Planning*, 22; 1;pp 65-72: New York
- [9] *Memorandum of Understanding.*, pasal 2 (1), hal. 3
- [10] Ogle, G.J., *Towards A formative Evaluations Tool*, (Dissertation: Virginia Polytechnic Institute and State University USA, 1982), 29.
- [11] Rossi, & Freeman, H. E., (1993). *Evaluation: A systematic Approach*, Newbury Park; Sage Publications
- [12] Sugiyono., (2006). *Metode penelitian pendidikan; kuatitatif, kualitatif dan R &D*, Bandung: Alfabeta
- [13] Stark, & Thomas., (1994). *Assessment and program evaluation*, Needham Heights: Simon & Schuster custom Publishing
- [14] Sukardi., (2000). *Evaluasi program pembelajaran*, Yogyakarta, Pustaka Pelajar
- [15] Valadez, J.& Bamberger, Ml. (2007). *Monitoring and Evaluating Social of perception of learning*, *Journal of European Industrial Training*; Vol. 31 no. 4
- [16] Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L.,(1996). *Programe Evaluation; Alternative approaches and practical guedelines 2nd Edition*, New York: Longman

THE INFLUENCE OF TEACHER PEDAGOGICAL COMPETENCE AND EMOTIONAL INTELLIGENCE TOWARDS MOTIVATION AND PHYSICS LEARNING RESULT OF STUDENT AT XI IPA GRADE SMA NEGERI 1 WATANSOPPENG

Kaharuddin Arafah¹, Adnani Yuni², Muris³

¹Lecturer at Phisycs Education Post Graduate Program, State University of Makassar

²Student at Phisycs Education Post Graduate Program, State University of Makassar

³Lecturer at Phisycs Education Post Graduate Program, State University of Makassar
ayuni2092@gmail.com, eltigakahar@yahoo.com

Abstract - This research is ex post facto which aims to determine the influence: i) the teacher pedagogical competence towards physics learning result; ii) emotional intelligence towards physics learning result; iii) the teacher pedagogical competence towards achievement motivation; iv) emotional intelligence towards achievement motivation; and v) achievement motivation towards physics learning result. The populations in this study were all students of XI IPA SMAN 1 Watansoppeng as many as 175 students. The process of collecting data using questionnaires and tests of physics learning outcomes that have been tested empirically. Data of the research result were by using analysis method of Structural Equation Modeling (SEM) with technique of Analysis of Moment Structural (AMOS). The analysis procedure is performed with descriptive analysis and inferential analysis, factor analysis and verification of structural model AMOS. Through the model can be concluded that: i) the teacher pedagogical competence doesn't have positive direct influence and significant toward physics learning result; ii) emotional intelligence doesn't have positive direct influence and significant toward physics learning result; iii) the teacher pedagogical competence has positive direct influence and significant toward achievement motivation; iv) emotional intelligence has positive direct influence and significant toward achievement motivation; and v) achievement motivation has positive direct influence and significant towards physics learning result.

Keywords: *eks post facto, pedagogic competence, emotional intelligence, achievement motivation, physics learning result*

I. INTRODUCTION

Education in all off country has a purpose which is to the intellectual life of the nation. So even with education in Indonesia. One indicator of the success of our education that increase learning result. In terms of improving learning outcomes it is necessary to examine the factors that could theoretically affect learning result. Had many previous studies that have found several factors either directly or indirectly to influence the improvement of learning result of students.

A lot of factors that can influence both internal factors and external factors. In terms of the factors of the learners most fundamental closely connected with the study results regarding its intelligence or intelligence factor, which relates to the behavior that plays a role in encouraging children to achieve, it is emotional intelligence.

Daniel Goleman in his book "Emotional Intelligence" write some indicators of emotional intelligence each building/give a positive relationship to the learning result, including self-motivated and able to build relationships with others. Learners are able to motivate themselves in terms of learning would make it viable to learn and believe with the results obtained. So even with the ability to build relationships with others, learners not being perfect so it must need help or cooperation with each other. In the case of learners who have a high EQ will be good at

building good relationships with friends so that they can complement each other/one another hang if there is thought to be less of each other. emotional intelligence has a relationship with the motivation and learning results.

There are many factors of the learner, the learning outcomes are also influenced by some external factors. The external factors that relate directly to the activities of learners that is a factor of teachers as learners, in this case the researchers took particular aspects of teacher competence pedagogic competence.

Teacher competence is needed in teaching and learning. In general, the teachers meet two categories, capability and loyalty. Capability is meant is that the teacher must have the ability in the field of science that teaches, has a theoretical capability of good teaching, from planning, implementation and evaluation. And that meant loyalty is loyalty teacher, which is loyal to the tasks not only in the classroom, but the pre-service or out servicing. Gilbert H. Hunt in his book "Effectively Teaching" (Rosyada, 2004).

Based on Government Regulation (PP) No. 16 of 2007 on teachers, there are many competencies that must be mastered by the teacher. One of them is the pedagogical competence. Under Article 28, paragraph 3 (a), pedagogic competence is the ability to manage learning of learners that includes an understanding of the learners, the design and implementation of learning, evaluation of learning outcomes, and the development of learners to actualize various potentials.

Pedagogical competence of teachers is absolutely necessary for successful learning and improving the quality of education. Without pedagogic, learning and education process will only road in place, there are no signs of improving the quality of education quality.

The success of student learning is part of the impact of teacher competence and skills in the learning process. The success of student learning are usually seen from the quality or changes shown students after participating in learning, so it can be assessed by the extent to which students' learning needs can be met optimally by the teacher. Thus the rationale encourage researchers to conduct research on the influence of teachers' pedagogical competence and emotional intelligence on achievement motivation and physics learning result of students at grade XI IPA SMAN I Watansoppeng. The problems in this research are: (1) Does emotional intelligence have positive direct influence towards achievement motivation? (2) Does emotional intelligence have positive direct influence towards physics learning result? (3) Does the teacher pedagogical competence have positive direct influence towards achievement motivation? (4) Does the teacher pedagogical competence have positive direct influence towards physics learning result? (5) Does the achievement motivation have positive direct influence towards physics learning result? The purpose of this study was to answer the formulation of the problem, namely (1) to determine the direct influence of emotional intelligence towards achievement motivation, (2) to determine the direct influence of emotional intelligence towards physics learning result, (3) to determine the direct influence of teacher pedagogical competence towards achievement motivation, (4) to determine the direct influence of teacher pedagogical competence towards physics learning result, and (5) to determine the direct influence of learning motivation towards physic learning result. The results of this study are expected to provide good benefits for students, teachers, schools and other researchers. For students, this study can provide input to further improve learning result. For teachers, this research as information can provide a positive discourse. For schools, this research can be used as information material in improving the teaching competence of teachers. As for other researchers, this study can be used as reference for further research, especially research that focuses on internal factors on students to subjects Physics.

II. RESEARCH METHOD

The type of research is the study "ex post facto", which is causality and correlation. This study attempted to investigate the direct influence of the independent variables namely the teacher pedagogical competence (X_2) and emotional intelligence (X_1) on physics learning result (Y_2) as dependent variables, both directly and through the achievement motivation (Y_1) as an intervening variable. Designs of linkages between these variables are described as follows.

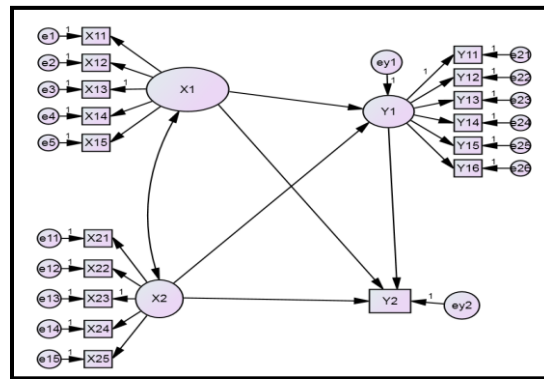


Figure 1. Structural Model of Functional Interwoven Proposed Among Variables Research

The populations in this study were all students of class XI IPA SMA Negeri 1 Watansoppeng academic year 2015/2016, Watansoppeng Regency, and South Sulawesi. Distribution of students in each class is shown in table 1 below.

Table 1. Distribution of Student Class XI IPA SMA Negeri 1 Watansoppeng

Class	Number of Student
XI IPA 1	33
XI IPA 2	35
XI IPA 3	34
XI IPA 4	34
XI IPA 5	35
Total	171

The technique used to determine the sample size is by slovin technique. As for the distribution of the sample is more presented in table 2 below.

Table 2. Distribution of Research Sample

Class	Sample
XI IPA 2	31
XI IPA 3	30
XI IPA 4	31
XI IPA 5	31
Total	123

The instrument used in this study consisted of questionnaire of the teacher pedagogical competence, emotional intelligence questionnaire, questionnaire achievement motivation, and test of physics learning result.

Before the instrument is ready for use, it must first be validated instrument that is validation of content and empirical validity. Contents validation test conducted on measuring instruments. Analysis of the contents of the instrument validation is done in this research that uses models Gregory in the form of a model agreement among experts. Empirical validation test performed on the test results the instrument consisting of item validity test and reliability test. Types and data collection techniques used in this study can be seen in the following table.

Table 3. Types and Data Collection Techniques

Data	Type of Data	Data Collection Techniques	Data Sources	Instrument
Teacher Pedagogical Competence	Interval	Questionnaire	Students	Questionnaire of Teacher Pedagogical Competence
Emotional Intelligence	Interval	Questionnaire	Students	Questionnaire of Emotional Intelligence
Achievement Motivation	Interval	Questionnaire	Students	Questionnaire of Achievement Motivation
Physics Learning Result	Interval	Questionnaire	Students	Questionnaire of Physics Learning Result

Analysis of the data used in this study consisted of a statistical analysis of descriptive and inferential analysis to test the hypothesis.

A. Analysis of Descriptive Statistic

A function of descriptive analysis is to provide an overview of the data obtained, such as: the number, maximum, minimum, mean, mode, median, standard deviation and variance.

B. Normality Test

Normality test is used to determine whether the data sample studied came from populations with normal distribution or not. Values were considered in determining test multivariate normality is critical ratio by using AMOS 22.0.

C. Linearity Test

Linearity test is performed to determine whether there is a linear relationship between the independent variables with the dependent variable using SPSS 22.0. The variable is said to be linear with other variables when $\text{sig.Linearity} < 0.05$.

D. Multicollinearity Test

Multicollinearity test is performed to determine whether there is a significant relationship between independent variables in a multiple linear regression model. Multicollinearity test was performed using statistical program SPSS 22.0 to see the value of Tolerance and Variance Inflation Factor (VIF) in the regression model.

E. Factor Analysis

Factor analysis was performed using AMOS 22.0 to be able to test the influence of the indicator with latent variables, a model must be eligible Goodness of Fit, which is an index that is used as a reference model is said to be acceptable fit. The index used is the Chi-square, CMIN / df, TLI, CFI and RMSEA.

III. RESULT AND DISCUSSION

A. Analysis Result of Research Data

1. Descriptive Statistic

Table 4. Summary of Descriptive Analysis Result

Statistic	Teacher Pedagogical Competence	Emotional Intelligence	Achievement Motivation	Physics Learning Result
Number of respondent	123	123	123	123
Number of item	34	33	40	26
Mean	122.9187	130.7073	133.8618	11.1382
Std. Error of Mean	1.09429	.94784	1.73110	.40838
Median	123.0000	131.0000	134.0000	11.0000
Mode	127.00	130.00 ^a	152.00	10.00
Std. Deviation	12.13624	10.51208	19.19884	4.52919
Variance	147.288	110.504	368.595	20.514
Skewness	-.020	-.548	-.163	.424
Std. Error of Skewness	.218	.218	.218	.218
Kurtosis	-.451	.868	-.370	.280
Std. Error of Kurtosis	.433	.433	.433	.433
Range	57.00	59.00	92.00	22.00

a. Multiple modes exist. The smallest value is shown

Research result data of variable the teacher pedagogical competence then presented the list of frequency distribution as shown in table 5 below.

Table 5. Distribution of Frequency, Percentage, and Categories for Teacher Pedagogical Competence

Interval Score	Category	Frequency	Percentage (%)
34,00 – 51,60	Very Low	0	0
51,70 – 78,50	Low	0	0
78,60 – 105,40	Moderate	3	2,44
105,50 – 132,30	High	67	54,47
132,40 – 170,00	Very High	53	43,09
Total		122	123

Research result data of variable the emotional intelligence then presented the list of frequency distribution as shown in table 6 below.

Table 6. Distribution of Frequency, Percentage, and Categories for Emotional Intelligence

Interval Score	Category	Frequency	Percentage (%)
33.00 - 59.80	Very Low	0	0
59.90 - 85.80	Low	0	0
85.90 - 111.80	Moderate	22	17,88
111.90 - 137.80	High	87	70,73
137.90 - 165.00	Very High	14	11,38
Total		123	100

Research result data of variable the achievement motivation then presented the list of frequency distribution as shown in table 7 below.

Table 7. Distribution of Frequency, Percentage, and Categories for Achievement Motivation

Interval Score	Category	Frequency	Percentage (%)
40,00 - 72,40	Very Low	0	0,00
72,50 - 104,40	Low	11	8,94
104,50 - 136,40	Moderate	56	45,53
136,50 - 168,40	High	52	42,28
168,50 - 200,00	Very High	4	3,25
Total		123	100

Research result data of variable physics learning result then presented the list of frequency distribution as shown in table 8 below.

Table 8. Distribution of frequency, Percentage, and Categories for Learning Result

Interval Score	Category	Frequency	Percentage (%)
0.0 - 5.6	Very Low	14	11.38
5.7 - 10.8	Low	44	35.77
10.9 - 16.0	Moderate	53	43.00
16.1 - 21.2	High	8	6.50
21.3 - 26.0	Very High	4	3.25
Total		123	100

B. Analysis Prerequisites Test

1. Data Normality Test

By using a significance level of 0.01, the data is said to be normally distributed if the critical ratio (cr) of kurtosis is between ± 2.58 . Based on the results of the output data normality test on Assessment of normality, normality test results obtained by the value of the multivariate cr kurtosis $2.262 < 2.58$ which means multivariate distribution is normal.

2. Linearity Test

Based on the results of linearity test the influences of X_1 with Y_2 have sig Linearity 0.000. For the influence of X_2 with Y_2 have sig Linearity 0.000. As for the influence of Y_1 with Y_2 have a sig Linearity 0.000. This means that the value of sig Linearity less than 0.05 ($0.000 < 0.05$). So it can be concluded that the influence of variable emotional intelligence with physics learning result, the influence teacher pedagogical competence with physics learning result and achievement motivation with physics learning result are linear and are eligible for further analysis.

3. Multicollinearity Test

Table 9. The Result of Multicollinearity Test

Model	Collinearity Statistics	
	Tolerance	VIF
Emotional Intelligence (X_1)	0.713	1.403
Teacher Pedagogical Competence (X_2)	0.794	1.260
Achievement Motivation (Y_1)	0.596	1.678

Dependent Variable: Physics Learning Result (Y_2)

C. Factor Analysis

1. Factor Analysis of Latent Variables

Test the fit between the theoretical models to empirical data can be seen at the level of Goodness of Fit Statistics.

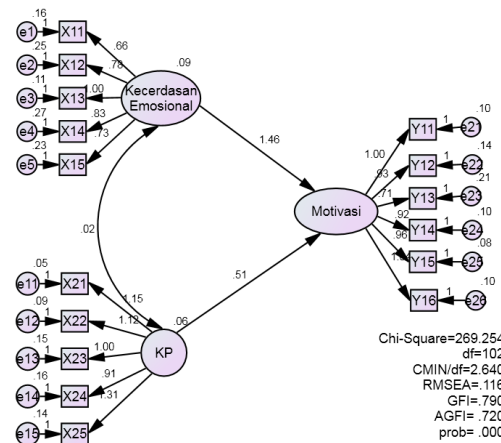


Figure 2. Initial Factor Model of Latent Variabels

The result of the factor analysis beginning shows that there are indices that do not meet the cut off value. Modifications made some errors influence of variables that have a value large change Chi-square. The modification result then re-analyzed by the results in Figure3.

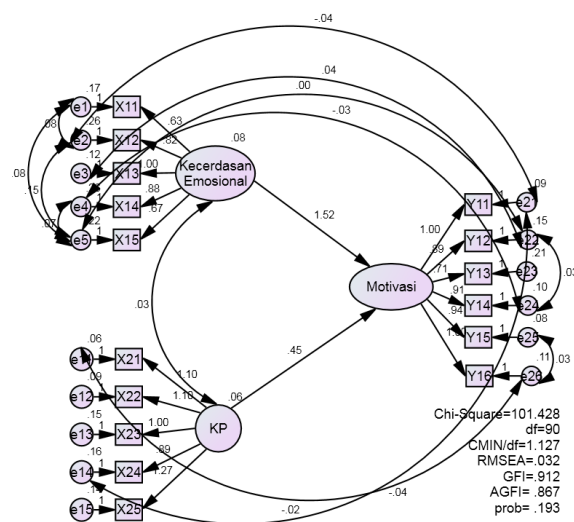


Figure 3. Final Factor Model of Latent Variabels

The final result can be seen that all the indexes have met the criteria so that these models can be received and analyzed further. Standardized regression weights can also show the influence of latent variables with the indicators.

Table 10. Standardized Regression Weights of Latent Variables

			Estimate
X11	<---	X1	.582
X12	<---	X1	.778
X13	<---	X1	.578
X14	<---	X1	.567
X15	<---	X1	.852
X21	<---	X2	.744
X22	<---	X2	.675
X23	<---	X2	.532
X24	<---	X2	.496
X25	<---	X2	.660
Y11	<---	Y1	.876
Y12	<---	Y1	.704
Y13	<---	Y1	.617
Y14	<---	Y1	.785
Y15	<---	Y1	.847
Y16	<---	Y1	.815

2. Model Verification and Final Model Development

Verified the theoretical model developed based on empirical data. An analysis of the Figure3 is the structural equation model of initial stages.

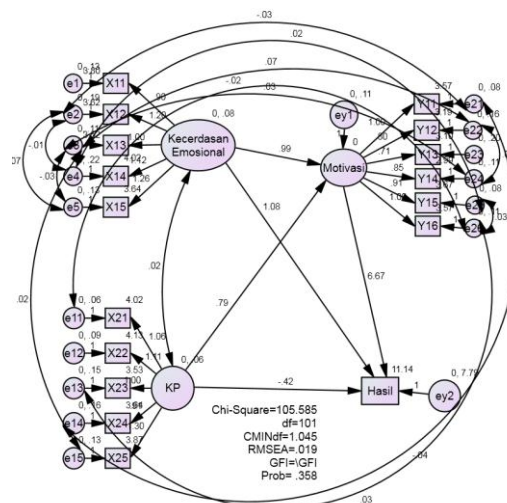


Figure 4. Structural Equation Model of Initial Stages

The results of the analysis of the initial stages in figure 4 shows that there are indices that do not meet the cut off value. Modifications made some errors influence of variables that have a value large change Chi-square. The modification result then re-analyzed by the results in Figure 5.

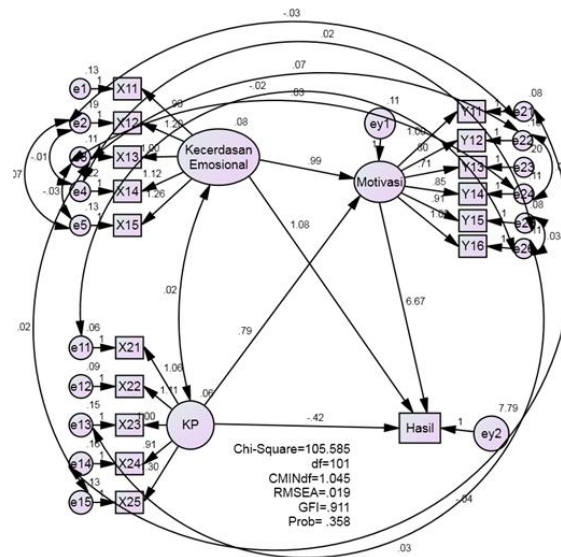


Figure 5. Structural Equation Model of Final Stages

In the final result can be seen that all the indexes have met the criteria so that these models can be received and analyzed further. Parameter of regression weighted shown in the table 10 below.

Table 11. Regression Weight of Final Model

			Estimate	S.E.	C.R.	P
Y1	<---	X1	.986	.186	5.290	***
Y1	<---	X2	.789	.212	3.726	***
Y2	<---	Y1	6.672	1.066	6.260	***
Y2	<---	X2	-.420	1.612	-.261	.794
Y2	<---	X1	1.084	1.567	.692	.489

Structural model is obtained based on the index overall fit can be seen in table 10. Mathematically structural equation model can be written:

$$\hat{Y} = 11.138 + 1.084X_1 - 0.42X_2 + 6.672Y_1 \quad R^2 = 66\%$$

D. Discussion

1. The Direct Influence of emotional Intelligence towards Achievement Motivation

For the fourth hypothesis testing showed that the influence variables of interpersonal intelligence towards learning motivation described in the regression weights of final model with the estimate results $\gamma_{x2y1} = 0.567$ with p value = $0.000 < 0.05$. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that emotional intelligence has positive direct influence and significant towards achievement motivation.

2. The Direct Influence of emotional Intelligence towards Physics Learning Result

For the second hypothesis testing showed that the influence variables of interpersonal intelligence towards physics learning result described in the regression weights of final model with the estimate results $\gamma_{x2y2} = 0.070$ with p value = $0.489 > 0.05$. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that emotional intelligence does not have positive direct influence and significant towards physics learning result.

3. The Direct Influence of Teacher Pedagogical Competence towards Achievement Motivation

For the third hypothesis testing showed that the influence variables of teacher professional competence towards learning motivation described in the regression weights of final model with the estimate results $\gamma_{x1y1} = 0.387$ with p value = $0.000 < 0.05$. This means that H_0 rejected and H_1

accepted at the significance level 0.05. This result indicates that teacher pedagogical competence has positive direct influence and significant towards achievement motivation.

4. The Direct Influence of Teacher Pedagogical Competence towards Physics Learning Result.

For the first hypothesis testing showed that the influence variables of teacher professional competence towards physics learning result described in the regression weights of final model with the estimate results $Y_{x1y2} = -0.023$ with p value = 0.794 > 0.05. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that the teacher pedagogical competence does not have positive direct influence and significant towards physics learning result.

5. The Direct Influence of Achievement Motivation towards Physics Learning Result

For the fifth hypothesis testing showed that the influence variables of learning motivation towards physics learning result described in the regression weights of final model with the estimate results $\beta_{y1y2} = 0.749$ with p value = 0.000 < 0.05. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that achievement motivation has positive direct influence and significant towards physics learning result.

IV. CONCLUSION AND SUGGESTION

The results showed that: 1) emotional intelligence has positive direct influence and significant towards achievement motivation; 2) emotional intelligence does not have positive direct influence and significant towards physics learning result; 3) the teacher pedagogical competence has positive direct and significant towards achievement motivation; 4) the teacher pedagogical competence does not have positive direct influence and significant towards physics learning result, and 5) the achievement motivation has positive direct influence and significant towards physics learning result.

Based on the research results obtained, it is advisable that matters; 1) For teachers to better understand the psychological factors that can affect the results of students; 2) for students to further develop the factors that was in him to get the optimal learning results; 3) For researchers interested in developing further this research, is expected to examine the limitations in this study, so further research can enhance this research.

REFERENCES

- [1] Goleman, D. (2004). *Emotional Intelligence (terjemahan)*. Jakarta: PT Gramedia Pustaka Utama.
- [2] Rosyada, D. (2004). *Paradigma Pendidikan Demokratis*. Jakarta: Prenada Media
- [3] Hamalik, O. (2001). *Proses Belajar Mengajar*. Jakarta: PT Bumi Aksara.
- [4] Undang-undang Nomor 16 Tahun 2007 Tentang Standar Kualifikasi Akademik dan Kompetensi Guru

EVALUATING POLICY IMPLEMENTATION INDICATORS IN DECENTRALIZED SCHOOLS

Lilik Sabdaningtyas¹, Budi Kadaryanto²

¹Department of Educational Sciences, University of Lampung

² Department of Language and Arts Education, University of Lampung
lilik_sabdaningtyas@yahoo.co.id

Abstract - Education decentralization in Indonesia has been initiated following the commencement of the reform era in the beginning of the new millennium which is resembled in a School Based Management (SBM) system. This brought impact on the schooling system as well, shifting the school management from the centralized into the decentralized one. The purpose of this study is to analyze the dynamic factors considered contributing to the performance of policy implementation of decentralization at the school level. In addition, it further evaluates those factors that serve as indicators for the schools performance in implementing decentralization at the school level. In a proportionally stratified random sampling, 182 respondents involved in the management of their school were asked to fill out the questionnaires about school climates, school policy executives, school policy targets, school resources as well as school organization management. This study was carried out in junior high schools in an urban area in Lampung Province. In a Structural Equation Modelling analysis, the result of the study are promising, as this indicates distinctive spectrum for evaluating the policy implementation for decentralized school management and its contributing indicators. These five indicators could be utilized as a means of evaluation how a school implements decentralization policies and as key performance indicators for their policy implementation. It further proposes a broader support for the reform among school principals, board members and related stakeholders to optimize their performance in implementing SBM.

Keywords: *education decentralization, evaluation, policy, School Based Management*

I. INTRODUCTION

In the beginning of the new millennium, the decentralization system in the government of Indonesia has brought the authority to school levels to manage and run their administrative activities in more democratic ways. This has been believed as an emerging phenomenon in most education systems in other countries [i][ii][iii]. Education decentralization has been associated with an act of giving more space for schools to improve student outcomes and the effectiveness of school systems by the transfer of authority. In other words, it was massively endeavored by practitioners, academicians, stakeholders, as well as schools to quest for the central government to devolve power and authority to even the lowest authority at the school level comprising school administrators, teachers, parents, as well as the community. At one hand, it is assumed that this situation brought euphoria for schools to develop and to run a democratic system, in which this had also existed in most countries during that period, those questing for the decentralization to occur at the school level had to be aware of the impact of turmoil of the new reform era turnover. This has been warned by Louis [iv] who suggests that educational reform (moving from centralized to decentralized one) is not a movement of no cost and difficulties as each level of schools has to develop their schools on their own. Louis further suggests that such reforms in education require extensive, consistent support, accompanied by in-service training and technical assistance for school leaders – enabling them to change management and planning skills, and helping them to deal with the school and classroom implications of reforms [v][vi][vii].

The implementation of School Based Management (SBM) in Indonesia has been always marked as educational reforms, not only driven by the in country turmoil as the fall of new order,

and entering the reform era, but this movement (SBM implementation) has actually been driven by many successful countries bringing a fundamental change in the school management and administration. SBM requires a school as the homefront of formal education strata to have a decision-making authority that stimulates and sustains improvements at the school. In the meantime, the degrees of school authority in making decisions in the domains of the school's mission, goals and school policies relating to financial, material and human resources should not simply delegated to them, rather transferring the authority to a representative managerial body called the school council or board (*Komite Sekolah*). With this way, there will be a synchronous actions within the school community, which may lead to have all parties' commitment to making school visions real under the umbrella of SBM [viii][ix][x][xi]

The government decision in succeeding the policy of SBM implementation has been partially driven by the belief that community control in the school management and administration will make schools become more effective and efficient in developing their visional programs. This idea came about in the United States in the mid-17th century. However, this idea had gained better and more widespread acceptance in 1980s, where the concept of community participation in SBM has become a major research themes and has been regarded as a policy initiative in school reforms in many education systems in the Europe and some parts of Asia. The Indonesian government itself was found a bit later in comparison to other Asian countries like Japan, China and Thailand in showing interest in implementing SBM [xii]. The idea of the SBM implementation should remain in public control while simultaneously fostering good instructional practices and good management tactics, including the prospect of effective accountability to all stakeholders. These decentralization initiatives take many forms, including the empowering of principals, teachers and parents. Today, educational decentralization with devolution of authority to individual institutions is a popular reform theme of governments around the world [xiii].

The government policy in making the transition from autocratic to democratic forms of government, has truly brought great impact in decentralizing educational systems in the country. This situation results in gaining citizen participation in government institutions. Winkler suggests that improving the quality of education is often offered as a goal of decentralization, reflecting the notion that local people can solve local education problems better than the centralized state system [xiv]. The success of SBM implementation will be the government challenge to assure the process run as it supposed to be. A policy will just be a dream that is stored as an archive alone if not implemented effectively. Thus the policy implementation is an important aspect that can determine the success or failure of a policy implementation in educational aspects [xv]. To determine the success or failure of policy implementation, it requires evaluative actions, as this will give portray of how the policy has been implemented. Once the implementation of policies that are not evaluated, then there will no information about the success of the implementation. This statement indicates the importance of the policy implementation evaluation itself and it crucial ties.

For the purpose of policy evaluation, it is required that proper evaluation model to evaluate the results to be accurate. Therefore, it is necessary to develop appropriate evaluation model to measure the performance of the implementation of the decentralization policy in the educational unit (schools). The purpose of this study is to propose an evaluation model to evaluate the performance of the implementation of the decentralization policy in the educational unit (schools). In a theoretical concept, achieving policy objectives (policy output) can be approached by actions that form the so-called resource mobilization policy input (input policy) and so-called resource management policy cycle (policy cycle at the stage of formulation, implementation, and policy evaluation [xvi]). This study focuses only on the stages of implementation of the policy (policy implementation). Furthermore, to produce a performance that effective policy implementation will require dynamic factors consisting of: environmental policy, implementing policy, group policy objectives, and resources, which is the fourth and policy input will be instrumental in the implementation of the policy cycle [xvii].

Based on the theoretical concepts, a model developed by referring to a model that uses a system of thought patterns and theoretical evaluation of policy implementation among a few researchers [xviii][xix][xx][xxi][xxii][xxiii]. They stated that the matching performance of policy implementation should be supported by a variety of factors covered by a policy system. Based on the opinion of experts, the factors supporting policy implementation performance in this study was developed by focusing on components of: input policy, cycle policy, and outputs policy. It is

necessary to develop appropriate evaluation model to measure the performance of the implementation of the decentralization policy in the educational units (schools).

II. METHOD

The design of this study is research & development (R & D) that utilizes a quantitative approach in a non-experimental design. This so-called evaluation models of policy implementation performance can be categorized as a study that utilizes a formal evaluation approach with a retrospective evaluation process. This kind of research tends to rely on the description of ex post facto (retrospective) about the ongoing program of activities, which in turn relates to the output as well as impacts of a policy implementation [xxiv]. Therefore, the design of such studies will not employ treatments or experiments by the researcher.

The sampling method used in this study was a proportional stratified random sampling technique. With this way, before selecting a random sampling technique, it should be initiated by determining the minimum number of the sample by using a minimum sample calculation formula of Cohen [xxv]. Based on Cohen's formula, power determined would be 60; with $\alpha = 0.01$; $R^2 Y.B = 0.10$; and $u = 4$. Therefore, the samples obtained in this study were at least 13 schools consisting of five state junior high schools and eight private junior high schools in Bandar Lampung.

The variables employed in this study consisted of both latent variables and observed variable. They were categorized as follows 1). School climate consisting of three observed variables; 2). School policy executives that consist of three observed variables. 3). School policy target consisting of also three observed variables. 4). School resources with four observed variables. 5). School organization management with three observed variables.

The data were collected by using questionnaires filled out on three groups of respondents namely the principal, teachers, and school committees as the policy executives at the school level as a result of decentralization in education. The instrument used consists of five items composed inventory each designed to get the data on school climates, school policy executives, school policy targets, school resources as well as school organization management, the target group of school policies, and school resources. A total of 115 items were developed by using Likert scale with four alternative answers. To determine the validity of the instrument, construct validity as well as content validity were used. The reliability of the instrument was calculated by using "Cronbach Alpha" with coefficient reliability of at least 0.5. Meanwhile, the data analysis was conducted in two phases namely: analysis of test requirements to test the normality and multikoleniaritas data, and test hypotheses to see the model fit using SEM with LISREL program [xxvi].

This research and development of model of current study adopted the model proposed by Cennamo & Kalk that was conducted in two phases namely: The first phase Instrument Development / Pre-Development of the steps which were carried out in the following steps; 1) theoretical development, 2) describing the theories into components, variables, and indicators in the form of table specification for the instrument development, 3) developing opaque instruments to be validated by using the Delphi technique policies, 4) developing instruments, as well as 5) testing the instruments. The test instruments were analyzed by using SPSS version 15.0. The test shall be imposed on the number of respondents from Junior High Schools 4 to 56 people consisting of 4 principals, 32 teachers, and 20 school committee. 6) The analysis of the results of testing instruments was carried out by using exploratory factor analysis (EFA). If the test turns out to be empirically valid and reliable, then the instrument developed can then subsequently be used as a tool to develop a theoretical model of the proposed research [xxvii].

The second stage of this research and development is development itself. The development of an evaluation of policy implementation was referring to model development steps proposed by Ferdinand [xxviii], Narimawati [xxix], and Ghozali [xxx]. The model development step employed in current study were: 1) theoretical/conceptualization of the model developed. 2) Developing of flow charts. 3) Converting the flow chart into a series of structural equation and measurement models specifications. 4) Selecting the estimation technique for the model to be created. 5) Product try out. The product try out was carried out to test the suitability of the products made by theoretical models with empirical models. Testing the suitability of the model imposed on 13 Junior High School respondents in Bandar Lampung that implement decentralization policies at the high school level education units comprising principals, teachers, and school committee. 6)

Evaluation of the model. Prior to the evaluation of the model, first evaluation requirements such models is the sample size, normality test, and test multicollinierity. The next step carried out was evaluating the model fit (goodness of fit) by using the alignment criteria. Alignment criteria used is, Chi-square (X^2) and probability (p), Root Mean Square Error of Approximation (RMSEA), normed Fit Index (NFI), Comparative Fit Index (CFI), Incremental Fit Index (IFI), Relative fit Index (RFI), Goodness of fit Index (GFI), Adjusted Goodness of fit Index (AGFI).

III. RESULTS AND DISCUSSION

A. Results

Overall, the instrument of performance of policy implementation indicated that there were no items that had communalities <0.45 , meaning that all items can be interpreted. Judging from the percentage cumulative for thirteen factors, the figure was found quite well with 77.59%. This percentage has met the requirements as stated by Nurosis^[xxxi] that when the percentage cumulative is more that 50%, it means that it is suitable decision factor. The Eigenvalue was found $1.142 > 1$, this figure shows that the factors raised can be used as an indicator of some traits. The reliability of the whole factor was found more than 0.70, meaning that all items can be declared as the reliable indicators.

In we take a look at the school organization management, the data show that the cumulative percentage was found good enough for the three factors which was equal to 69.32%. This resembles that the instrument of school organization management could explain the theory of three-dimensional factor of as much as 69.32% with the Eigenvalue reaching to $1.286 > 1$. This figure shows that the factors raised can be used as an indicator of the traits. Meanwhile, the common factor was found to have no items smaller than 0.45, meaning that all items can be interpreted. Further, the reliability of the whole factor was also found more than 0.70, meaning that all items can also be declared as the reliable indicators.

Subsequently, the result of the school climate instruments shows that the common factor (communalities) was found to have no items smaller than 0.45 meaning that all items can be interpreted. Judging from the cumulative percentage, the finding was good enough for the three factors with the percentage of 76.34%. This means that the instrument for evaluating the policy implementation about the school climate could explain the dimensional theory of three factors as much as 76.34% and the Eigenvalue was found to be $1.462 > 1$. This figure shows that the factors proposed in the instruments can be used as indicators of the traits. The reliability value of the whole factors was found more than 0.70, meaning that all items can be declared as the reliable indicators as well.

In the school policy executives' factor, the data show that the common factor was found to have no items smaller than 0.45 meaning that all items can be interpreted. Judging from the cumulative percentage, the figure was found to be good enough for the three factors which was equal to 85.08%, making the instrument of school policy executives can explain the theory of three-dimensional factor of as much as 85.08% with the Eigenvalue for $1.045 > 1$. This figure shows that the factors raised can be used as the indicator of the traits. The reliability of the whole factor was also found to be more than 0.70, meaning that all items can be declared as the reliable indicators as well.

In the school policy target, experimental data show that the common factor was found to have no items smaller than 0.45 meaning that all items can be interpreted very well. The cumulative percentage was also found good enough for the three factors which was equal to 76.52%, and making the instrument for the school policy target can explain the theory of three-dimensional factor of as much as 76.52% with the Eigenvalue of $1.045 > 1$. This figure shows that the factors proposed can be used as an indicator of the traits. The reliability of the whole factor was also found to be more than 0.70, meaning that all items can be declared as the reliable indicators as well.

In the meantime, the school resources data show that the common factor indicated that there was only one item whose value is smaller than 0.45 with the value of 0.397. Other items had common factor value to be more than 0.45, meaning that most of the items can be interpreted. Judging from the cumulative percentage, it was also found good enough for the four factors that was equal to 66.98%. This means that the instrument can explain the dimensional theory of the four factors which were found to be more than 66.98% with the Eigenvalue $1.053 > 1$. This figure shows that the factors raised can be used as an indicator of the traits. The reliability of the whole

factor was also found to be more than 0.70, meaning that all items can be declared as the reliable indicators as well.

The data normality test results was found to have abnormal distribution for the manifest y9, Y11, x10 and x11. This is characterized by a chi-square value with a probability of less than 0.05. About the manifest of the abnormal distribution, normalization of the data was carried out [xxxii]. After the normalization test, it shows all manifests have chi-square value with a probability of more than 0.05, then it's normal. Normality test results are all manifests; Y4 = 0.078; Y5 = 0.118; Y6 = 0.122; Y7 = 0.090; Y8 = 0.055; Y9 = 0.997; Y10 = 0.074; Y11 = 1.000; Y12 = 0.104; Y13 = 0.110; Y14 = 0.087; Y15 = 0.109; Y1 = 0.140; Y2 = 0.075; Y3 = 0.117; X1 = 0.175; X2 = 0.067; X3 = 0.082; X4 = 0.077; X5 = 0.098; X6 = 0.200; X7 = 0.096; X8 = 0.162; X9 = 0.080; X10 = 0.998; X11 = 0.999; X12 = 0.375; X13 = 0.067.

The multicollinearity test was intended to see whether there is a perfect correlation or large value among the independent variables. If there is a magnitude of correlation coefficients more than 0.80, it means that the multicollinearity is present. Based on the test results of the data multicollinearity, it turns out that the correlation among the dependent variables was nothing found more than 0.80. This means that there is no multicollinearity of the data.

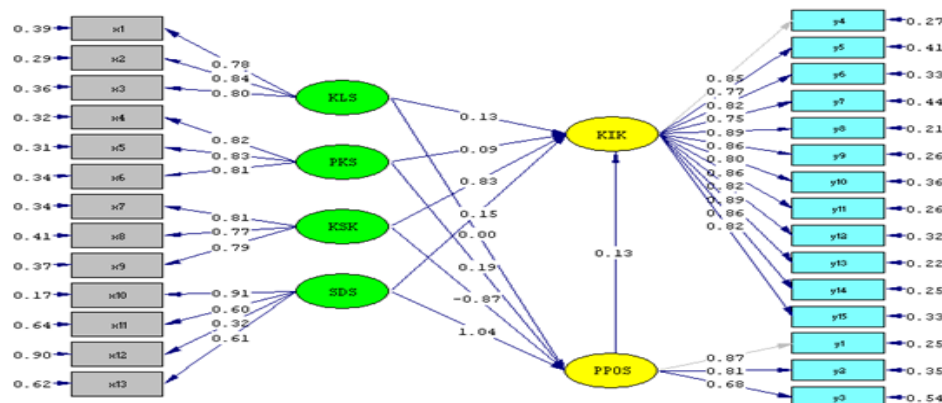


Figure 1. The evaluation of policy implementation model [xxxiii]

The evaluation of policy implementation model can be seen in Fig. 1. The test results of the proposed model has a goodness of fit index parameter which does not meet the entry requirements model of the chi-square = 534.2862 (not met), $p = 0.000$ (not met), and RMSEA = 0.06149 (not met). Then later model was modified by correlating between the two errors indicators that have large residual covariance. The result of these modifications successfully reduced chi-square and probability values, so that the model becomes suitable. Fig.2 is modeled after the modification. Apparently the result of the calculation of the Goodness of Fit Index has met the criteria. Calculation of the model is that the modified chi-square = 370.048 (met), $p = 0.07363$ (met), and RMSEA = 0.031 (met). The results of this study can be interpreted that the model developed is received after modified.

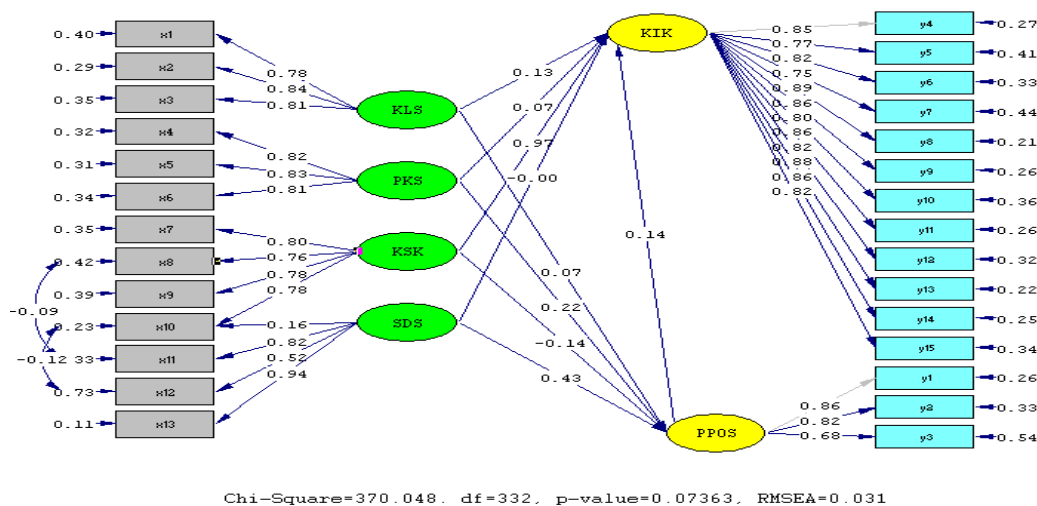


Figure 2. The evaluation of policy implementation model [xxxiv]

B. Discussions

These findings support the theory of policy implementation model proposed by several previous studies such as by Grindle [xxxv] which states that any implementation of the policy needs to consider the context or environment (school climate) in which the action (action) is performed. Referring to the opinion of Grindle, it means that school climate is an important determinant of performance policy implementation. Mazmanian and Sabatier argued that the implementation of the policy will be effective when the implementation comply with what has been laid down by the regulations (technical instructions, implementation guidelines). Further, they stated that the target group must also comply with the program, regardless of their compliance with the policy objectives whether or not achieved. Referring to the opinion of Mazmanian and Sabatier, it means that the target groups and implementing policies is also an important determinant of performance policy implementation [xxxvi].

Edwards argues that one of the variables that affect policy implementation is resources (resource). In this case necessary for the implementation of a policy is not only about the staff number / personnel sufficient but preferably expertise or skills in implementing policies that have been formulated. Edwards refers to the opinion that the quality and quantity of resources, especially human resources is a factor supporting the implementation of the policy [xxxvii]. Furthermore, according to Smith there are four important components in the implementation of the policy; the idealized policy, the target group, the implementing organization, environment factors. On the basis of Smith's opinion that the target group policy, implementing policy and environmental factors are also factors supporting the policy implementation [xxxviii].

Meter & Horn suggested that the performance of the implementation of the policy can also be determined by factors; standards and targets, resources, communication between organizations, the characteristics of the implementing organization, and the social, economic and political [xxxix]. These five factors will shape attitudes towards implementing policies that will be implemented and ultimately affect the performance of the policy implementation. Cheema & Rondinelly suggests four factors that are thought to influence the implementation of decentralization policies suggested as the independent variable; environmental conditions, the relationship between the organization, resources, and the character of the executing agent [xl]. Implications of the acceptance of this theoretical model is expected to be followed-up for the program manager to conduct training on the performance of schools in implementing the decentralization policy of compulsory education schools that implement the functionality.

The policy implementation for SBM in Indonesia has been largely educational reforms following the reform era. This movement (SBM policy implementation) has actually been driven by many successful countries bringing a fundamental change in the school management and administration. SBM requires a school as the homefront of formal education strata to have a decision-making authority that stimulates and sustains improvements at the school. In the meantime, the degrees of school authority in making decisions in the domains of the school's

mission, goals and school policies relating to financial, material and human resources should not simply delegated to them, rather transferring the authority to a representative managerial body called the school council or board (*Komite Sekolah*). With this way, there will be a synchronous actions within the school community, which may lead to have all parties' commitment to making school visions real under the umbrella of SBM. This has been regarded as the the foundation of the SBM policy implementation success. One of which way has been proposed to be evaluated by utilizing the policy implementation performance indicators discussed above

IV. CONCLUSION

The current study is trying to propose an evaluation model that evaluates the performance of a policy implementation in the context of education decentralization at the school level. The evaluation of policy objectives achievement (policy output) can be approached by actions, either in the input policy level, policy implementation, or the management policy level (formulation, implementation, and policy evaluation). This study focuses on the stages of implementation of the policy (policy implementation). School climates, school policy executives, school policy targets, school resources as well as school organization management were found applicable to be utilized for evaluating the policy implementation for decentralized school management and its contributing indicators. It gives implication for a broader support for the reform among school principals, board members and related stakeholders to optimize their performance in implementing SBM.

REFERENCES

- [ⁱ] Patrinos, H. A., & Fasih, T. (2009). *Decentralized decision-making in schools: The theory and evidence on school-based management*. World Bank Publications.
- [ⁱⁱ] Rado, P. (2010). Governing decentralized education systems: Systemic change in South Eastern Europe.
- [ⁱⁱⁱ] Gershberg, A. I., González, P. A., & Meade, B. (2012). Understanding and improving accountability in education: A conceptual framework and guideposts from three decentralization reform experiences in Latin America. *World Development*, 40(5), 1024-1041.
- [^{iv}] Louis, K. S. (1998). "A light feeling of Chaos": Educational reform and policy in the United States. *Daedalus*, 127(4), 13-40.
- [^v] Woods, D. (2002). Moving Forward... From Where You Are to School Improvement That Lasts: A Research-Based Guide.
- [^{vi}] Fiszbein, A. (Ed.). (2001). *Decentralizing education in transition societies: Case studies from Central and Eastern Europe*. World Bank Publications.
- [^{vii}] Gamage, D., & Sooksomchitra, P. (2006). Decentralisation and school-based management in Thailand. In *Decentralisation and Privatisation in Education* (pp. 151-167). Springer Netherlands.
- [^{viii}] Caldwell, B. J. (2005). *School-based management* (Vol. 3). Paris: International Institute for Educational Planning.
- [^{ix}] Sumintono, B. (2006). Decentralized centralism: School based management policies and practices at state secondary schools in Mataram, Lombok, Indonesia.
- [^x] Patrinos, H. A., & Fasih, T. (2009). *Decentralized decision-making in schools: The theory and evidence on school-based management*. World Bank Publications.
- [^{xi}] Bandur, A. (2009). The implementation of School-Based Management in Indonesia: Creating conflicts in regional levels. *Journal of NTT Studies* Vol, 1(1), 17.
- [^{xii}] Shoraku, A. (2009). Educational movement toward school-based management in East Asia: Cambodia, Indonesia and Thailand. *Background Paper for EFA Global Monitoring Report*.
- [^{xiii}] Hanson, E. M. (1998). Strategies of educational decentralization: Key questions and core issues. *Journal of educational administration*, 36(2), 111-128.
- [^{xiv}] Winkler, D. R. (1993). Fiscal decentralization and accountability in education: Experiences in four countries. *J. Hannaway y M. Carnoy (Comps.), Decentralization and school improvement: Can we fulfill the promise*, 102-134.
- [^{xv}] Udoji, C. J. (1981). The African public servant as a public policy maker. *Public Policy in Africa, Africa Association for Public Administration and Management*. Addis Abeba.
- [^{xvi}] Dunn, W. N. (2015). *Public policy analysis*. Routledge.
- [^{xvii}] Mustopadidjaja, AR. (2002). Manajemen proses kebijakan publik. formulasi, implementasi, dan evaluasi kinerja. Jakarta: Lembaga Administrasi Negara.
- [^{xviii}] Grindle, M. S. (Ed.). (1980). *Politics and policy implementation in the Third World* (Vol. 310). Princeton, NJ: Princeton University Press.
- [^{xix}] Mazmanian, Daniel., & Sabatier, Paul. (1986). Effective policy implementation. Lexington: D.C, Heath
- [^{xx}] Ingram, H. M., & Mann, D. E. (Eds.). (1980). *Why policies succeed or fail* (Vol. 8). SAGE Publications, Incorporated.
- [^{xxi}] [38] Smith, T. B. (1973). The policy implementation process. *Policy Sciences*, 4(2), 197-209.
- [^{xxii}] Van Meter, D. S., & Van Horn, C. E. (1975). The policy implementation process a conceptual framework. *Administration & Society*, 6(4), 445-488.
- [^{xxiii}] Rondinelli, D. A., Nellis, J. R., & Cheema, G. S. (1983). Decentralization in developing countries. *world bank staff working paper*, 581.
- [^{xxiv}] Dunn, William N. (1998). Pengantar analisis kebijakan publik (ed II). cetakan I. (terjemahan Samodra Wibawa, dkk). Yogyakarta: UGM Press. (Buku aslinya diterbitkan tahun 1994).

- [^{xxv}] Cohen Jacob. (1977). *Statistical Power Analysis for the behavioral Sciences*. New York San Francisco London: Academic Press, A Subsidiary of harcourt brace Jovanovich, Publishers
- [^{xxvi}] [28] Ferdinad, A. 2000. *Structural Equation Modeling Dalam Penelitian*, Badan Penerbit Universitas Diponegoro, Semarang
- [^{xxvii}] Cennamo, K., & Kalk, D. (2005). The professional designer. *Real world instructional design*, 272-285.
- [^{xxix}] Narimawati, U. (2007). *Riset Manajemen Sumber Daya Manusia*. Jakarta: Agung Media.
- [^{xxx}] Ghozali, I. (2005). *Model Persamaan Struktural*. Semarang: UNDIP.
- [^{xxxi}] Nurosis, M. J. (1993). *SPSS. Statistical data analysis*. Spss Inc.
- [^{xxxii}] Setyo Heri Wijayanto. (2008). *Structural equation model*. Yogyakarta: Graha Ilmu.
- [^{xxxiii}] [34] Lilik Sabdaningtyas. (2012). *Model Teoritik Evaluasi Kinerja Implementasi Kebijakan*. Makalah Proceeding Konferensi Ilmiah Nasional. Surabaya: Kampus Unesa.
- [^{xxxv}] Grindle, M. S. (Ed.). (1980). *Politics and policy implementation in the Third World* (Vol. 310). Princeton, NJ: Princeton University Press.
- [^{xxxvi}] Mazmanian, Daniel., & Sabatier, Paul. (1986). *Effective policy implementation*. Lexington: D.C, Heath
- [^{xxxvii}] Edward III., George C (1980). *Implementing public policy, congressional quarterly*. Washington, D.C. Ferdinand, Augusty. (2000). *Structural equation modeling dalam penelitian manajemen. aplikasi model-model rumit dalam penelitian untuk tesis S2 dan disertasi S3*. Semarang: Badan Penerbit Universitas Diponegoro.
- [^{xxxix}] Meter, D.S., & Horn, C.E. (1973). *The policy implementation process a conceptual framework in administration and society*. Beverly Hills: Sage Publication.
- [^{xl}] Rondinelli, D. A., Nellis, J. R., & Cheema, G. S. (1983). *Decentralization in developing countries. world bank staff working paper*, 581.

IDENTIFICATION CRITICAL THINKING SKILLS OF SMA MUHAMMADIYAH 1 BANJARMASIN STUDENTS TO THE MATTER DYNAMIC ELECTRICITY

Misbah¹, Saiyidah Mahtari², Sayid Muhammad Hasan³

^{1,2,3} Physics Education Study Program Lambung Mangkurat University, Banjarmasin, Indonesia
misbah_pfis @ unlam.ac.id

Abstract - The study is done to identify the critical thinking skills of students in senior high school. This study is aimed to describe the critical thinking skills students covering five skills, standard clarification, standard endorsement, conclusion (inference), advanced clarification, and strategies and tactics. The research is descriptive quantitative research by giving test critical thinking skills to students of multiple choice. The research was conducted in SMA Muhammadiyah 1 Banjarmasin. The subject research was tenth grade students at SMA 1 Banjarmasin in the academic year 2015/2016. The researcher took 30 students as the sample. The result showed critical thinking skills that students at elementary clarification by 95.83%, basic suport by 56.67%, 76.67% by inference, advanced clarification 3.33%, and strategies and tactics of 95.83%. The result shows that students have a good score on standard clarification, clonclusion (inference) and strategies and tactics, but still low on standard clarification and advanced clarification.

Keywords: *identification, critical thinking skills.*

I. INTRODUCTION

The essentially of physic is a collection of knowledge, ways of thinking, and investigation. Physics is seen as a process and a product (Prasetyo, 2004: 31)^[1]. Physics on recited the objects of the research in the form of things and natural events is using standard procedures that is called the method or scientific process. Therefore, the learning process and learning outcomes assessment physics should reflect the scientific characteristics (Mundilarto, 2010: 4)^[2].

According to Ennis (Costa, 1985: 54)^[3], critical thinking is rational and reflective thinking that is focused on what is believed and done. Rational means having faith and a view which supported by the evidence standard, actual, fairly, and relevant. Reflective means considering actively, diligently and carefully of all the alternatives before making a decision. It means that critical thinking requires the use of multiple strategies to produce a decision as a basis for taking action or belief.

Furthermore, Scriven (Fisher 2009: 10)^[4] argues that critical thinking is the interpretation and evaluation of skilled and active versus observation and communication, information and arguments. Critical thinking teaches learners to think about how they make a conclusion, defend positions on complex issues, consider the various kinds of viewpoints, analyze concepts, theories and explanations, explaining the issues and conclusions, solve problems, move the idea to the new context, examine assumptions, assessing the facts, and explore the implications and consequences, and getting to know the contradictions and inconsistencies in their own thoughts and experiences (Paul & Elder, 2007: 8)^[5].

Ennis and Noris (Nitko, 2008: 74)^[6] is divided the components of the ability to master knowledge into five skill, hereinafter referred to as critical thinking skills, those are:

1. Elementary clarification, include: focusing on the questions, analyzing the arguments, asking and answering questions that require clarification or challenge.

2. Basic support, include: considering the credibility of the resource and deliberation of observation.
3. Inference, include: conducting and considering the deduction, induction, the decision of values.
4. Advanced clarification, include: identifying the technical term, considering the definition, and identifying the assumption.
5. Strategies and tactics, include: determining an action and interacting with the others.

Based on the statements above, it can be concluded that the critical thinking skills are a high level thinking skills by using of knowledge to take an action or decision. With the critical thinking students are trained to connect the concepts to new situations, making inferences, analyzing, evaluating, and implementing a strategy. In this study, the indicator of critical thinking skills can be seen in the Table 1 below:

Table 1. The Indicator Of Critical Thinking Skills

Critical Thinking	Basic of Critical Thinking	Explanation
Standard Clarification	Analyzing the arguments	To find the similarities and the differences
Standard Endorsement	Considering the credibility a resource	To capable giving a reason
Conclusion (inference)	Making the deduction and considering the induction	To make a generalisation To make a conclusion and hypothesis
	making and considering the decision of values	The application of principles
Advanced Clarification	Defining a technical term	Definitions' strategy: action, measure, and identification
Strategies dan tactics	Deciding an action	To formulate the possibles' alternatives

II. RESEARCH METHODS

The type of this study is descriptive research. The data of research is conducted in class X SMA Muhammadiyah 1 Banjarmasin with 30 students of the total population. The purpose of the study was to describe the critical thinking skills of students including five skills, those are the standard clarification, standard endorsement, conclusion, advanced clarification, and the strategies and tactics. The research is descriptive quantitative by giving tests of critical thinking skills to students. The test method was used to obtain the data of the students ability in critical thinking skills to solve the physic examinations of dynamic electricity. This study used the multiple choice test. The students did the test in 90 minutes, after that the researcher did the analysis of the students answers. The results of the analysis are used to describe the thinking skills of students. To gain the standard of mastery score criterion - *Kriteria Ketuntasan Minimal* (KKM) of indicator, the following formula is used:

$$\text{KKM of indicator} = \frac{\sum \text{indicator score achieved by students}}{\sum \text{maximum score}} \times 100\%$$

III. RESULT AND DISCUSSION

The results of 30 students who did 25 test item of critical thinking skills are described in the following table:

Table 2. The Result Critical Thinking Skills

No	Critical thinking skills	KKM of Indicator
1	Standard clarification	95.83%
2	Standard endorsement	56.67%
3	Conclusion	76.67%
4	Advanced clarification	3.33%
5	Strategies and Tactics	95.83%

The Thinking skills is an important thing that must be held by students for the development of science and technology nowadays. One of the purpose of learning physics is able to solve

problems related to life activities in everyday. Solving the problem was required a high level of critical thinking skills. The critical thinking skills of a student would be helpful to make an appropriate decisions, accurate, systematic, correct and logical to consider the different of perspectives or aspects.

The estimate of critical thinking skills is held after the students have learned the materials about dynamic electricity. The students in standard Clarification and the strategies and tactic got the highest indicator of the KKM. The students in the standard clarification that included to analyze the arguments, this aspect got a mastery score of 95.83%. The result showed that students are able to analyze the arguments. In the aspect of strategies and tactics that include determine a course of action, this aspect got a mastery score of 95.83%. The result showed that students are able to use the aspect of strategies and tactics. The lowest score of critical aspect of is in the advanced clarification aspect, this aspect got a score of 76.67%. Te result was indicated that most of the students were not able to make an induction and consider the induction. In the aspect of standard endorsement is concluded to consider the credibility of a source of students, they only got 56.67% of mastery score of indicator. The lowest of aspect of critical thinking skills was in advanced clarification, the aspect got 3.33% of mastery score (KKM) of indicator.

Students are still not able to provide in the standard clarification dan in advavced clarification. So students have not been able to give a reason in considering the credibility of a resource and in the aspect of dentifications aspects. The critical thinking skills of students was not reached because the learning that has not been aimed at improving critical thinking. To improve critical thinking skills, can be trained continuously. Because only with exercise, the students can make the critical thinking skills becomes a habit. Critical thinking is a habit of thinking that should be instilled from an early age. And everyone has the ability to be critical thinkers reliable. Critical thinking can help anyone to be understand how to observe themselves, how the students see the world, and how it relates to the others, helping good attitude and the behavior, and evaluating themselves. The critical thinking allows anyone to analyze they own thoughts to ensure that they have the choice and making the best intelligent best conclusions. Meanwhile, people who do not think critically, they can not decide for themself what they thinking about, what they believe and how to act. Because of having failed to think independently, they would imitate the others, adopted their ideology and accepting the conclusions of others passively.

Student-centered learning allows the discussion. Discussions is one effective way to train and develop the critical thinking skills of air, because: (1) through discussion, students share opinions, thinking perspectives, and gain the experience; (2) through discussion students may consider, reject or accept its own opinion as well as opinions of other students to conform to answer or opinion of the group; and (3) through discussions also the students can make adjustments or reduce barriers between himself and another student so that he is free to think and act. Interaction among students, students and teachers made in the discussions is very influential on the growth and development of students' critical thinking dispositions (Lambertus, 2009)[7]. The thingking skills will be trained continuously (continuous) until it become sa habit, so when students are in the problem, they can make decisions quickly, accurately, and efficiently. The thinking skills is becoming a provision for students to compete in the era of globalization.

IV. CONCLUSION

This results of the data indicated that students have a good score in standard clarification, conclusion (inference) and strategy and tactics aspects, but still low on standard endoresement and advanced clarification. The aspects of critical thinking skills to be uneven mastered by students. The critical thinking skills must be trained continuously, until it becomes a habit. This habit will be the basic attitude, and ultimately formed the disposition of critical thinking.

ACKNOWLEDGMENT

All praise and gratitude to Allah SWT researchers, because thanks to His grace and guidance we can finish the study. Authors thank to headmaster SMA Muhammadiyah 1 Banjarmasin for research permission at the school.

REFERENCES

- [1] Prasetyo, Z. K. (2004). *Selekta capita of learning physics*. The second edition. Jakarta: The Open University.
- [2] Mundilarto. (2010). *Physics learning outcomes assessment*. Yogyakarta: P2IS JURDIK Physics, State UNY.
- [3] Costa, A. L. (ed.) (1985). *Developing minds: A Resource Book for Teaching Thinking*. Alexandria Virginia: Association for Supervision and Curriculum Development.

- [4] Fisher, A. (2009). Critical thinking: an introduction. Jakarta: Erland.
- [5] Paul, R., & Elder, L. (2007). Critical thinking standards. The foundation for critical thinking press.
- [6] Nitko, A. J., & Brookhart, S. M. (2008) .Assessment and grading in classrooms. Upper Saddle River: Merrill / Prentice Hall.
- [7] Lambertus. 2009. Importance of Training Critical thinking skills in Learning Mathematics in elementary school. Education Forum, Volume 28, No. 2, 136-142

The Influence of the Socio-Cultural-Based Learning Device to Student Academic Performance

Muhammad Nur Wangid, Ali Mustadi
Yogyakarta State University

Abstract- This article is part of research on the development of socio-cultural-based thematic integrative learning device. One concern in the education process in general is that a student's academic performance is not satisfactory. To overcome these problems the development of devices that allow students to interact with their social environment to the fullest is made. The learning materials available in the environment is packed in a single comprehensive learning device. The device allows the learning happens naturally, so that students learn naturally because they are in the nuances of social context that is already very familiar with the material substance. To see the effectiveness of these devices, experiments were done. The results showed that students who experience learning device based on socio-cultural learning appears to have a different academic performance significantly compared to the academic performance of students who experience learning device prepared by teachers.

Keywords: *Socio-Cultural-Based Learning Device; Academic Performance*

I. INTRODUCTION

Curriculum development undertaken by the government is one of the efforts in order to overcome the problems that hit the nation. One of them is the character values of the nation begins to decline. Including socio-cultural values are marginalized by the values and foreign cultures. Such conditions also experienced by elementary school in Yogyakarta. Therefore, it is necessary to develop sociocultural-based thematic integrative learning model at elementary schools in the region. Based on those things implies that the process of formal environmental education in particular must not be separated from social and cultural values that exist in the environment. Thus, the socio-cultural values that have been preserved by our predecessors did not fade as the time being. Therefore an effective sociocultural-based learning device needs to be developed by teachers so that learning is more meaningful for students.

According to Kyriacou (2009: 7), "effective teaching can be defined as teaching that successfully achieves the learning by pupils intended by the teacher". The learning success is based on the development of the effective learning devices to deliver students toward learning goals. Therefore, the learning device is one of the important things that need to be considered in the implementation of learning in schools. The good learning device will support learning activities in the classroom in accordance with the expected goals. Teachers need to design a good learning device that enable learning in the classroom to be enjoyable. Learning in the classroom that is well designed by teachers support the success of the learning. Furthermore Suhadi (2007: 3) revealed that the learning device was a number of materials, equipment, learning media, instructions and guidelines that will be used in learning activities.

In elementary school the learning that use thematic-integrative approach that is an integrative learning use an approach across subject areas. This approach is done by combining the fields of study in one theme by setting curricular priorities and find the skills, concepts and attitudes that overlap in several areas of study. In this thematic-integrative learning, themes related and overlapping each other is the last thing teachers want to search and select in the planning stage of the program. Thematic-integrative learning occurs when authentic events or exploration of the topic/theme become controller in learning activities, so that students learn the process and content of some subjects simultaneously.

Thematic-integrative learning will go well if it is supported by the media or a real object using visual symbols. Cognitive learning outcomes of learners gained from direct experience (concrete), the reality on the life environment of learners then through mock objects, until the

verbal symbols (abstract). As is the case according to Dale, the more upper at the top of the cone the more abstract the learning media conveys that message. This sequence does not mean the process of learning and teaching interaction on thematic-integrative learning must always be started from direct experience, but begins with the kind of experience that best suits the needs and abilities of a group of learners faced with situations of learning and the surrounding environment.

Sociocultural-based learning is learning that integrate sociocultural values in the learning process. Sociocultural values in question are cultural values in the surrounding environment of the learners. Sutarno (2007: 7-2) stated that the use of local cultural (ethnic) in learning is very useful for meaning and the cognitive learning process, because students gain firsthand experience that is contextual (footbridge mine) and apperception materials to understand the concept of science in local cultural (ethnic). From these statements, it can be observed that, by sociocultural-based learning will be more meaningful because of the things that are taught by teachers are contextual to the values of the surrounding culture..

From the view of Vygotsky (Schunk, 2012: 243) it can be interpreted that the sociocultural theory, the learning process experienced by learners can not be separated from the social interaction and the use of cultural values and symbols surrounding environment. In line with the conclusions of Higher Education (Sutarno, 2007: 7-2) mentioned that the sociocultural-based learning cultural values are integrated as a tool for learning to motivate learners to apply knowledge, work cooperatively, and perceived linkages between various subjects. From these statements can be observed that the socio-cultural-based learning is very relevant to be applied in order to create a learning environment that is meaningful and enjoyable for the learners. Further described by Sutarno (2007: 7-6) that the culture-based learning can be divided into four kinds, namely learn about the culture, learn with the culture, learn through culture and learn to have a culture. Following the integration of the subject matter in the context of socio-cultural student:

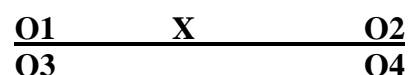
Sub Theme	Learn through culture (Media)	Learn with the culture (Outcomes)	Learn to have a culture (lesson plans or Rencana Program Pembelajaran; RPP)
My experience with friends	<ul style="list-style-type: none"> Gamelan, Angklung 	<ul style="list-style-type: none"> Delikan Games Clay Flour Games 	<ul style="list-style-type: none"> Imitate the discipline, honesty and characterization of the characters in the story.

To carry out integrative thematic learning the following stages is executed. The planning stage which includes the selection of themes, determine the type of subjects, choosing study material, Competency Standards (Standar Kompetensi, SK), Basic Competency (Kompetensi Dasar, KD) and Indicators, specify sub-skills are combined, formulate indicators of cognitive learning outcomes, determining the learning steps; then the implementation stage; and the last one is evaluation stage.

Based on the description above, then the following research hypothesis is proposed. "There are significant differences concerning the increase of cognitive learning outcomes among students who follow learning by using socio-cultural-based learning devices than the increase in cognitive learning outcomes of students who take the conventional learning".

II. METHODS

The type of research used in this study is a quasi-experimental with design as follows.



Explanation: O1 : Experiment class pretest
 O3 : Control class pretest
 X : Given Treatment
 O2 : Experiment class posttest
 O4 : Experiment class posttest

The experimental group was given treatment in the form of learning using socio-cultural-based learning devices that has been declared fit or validated by experts in the field (Nur Wangid

et al, 2014). While in the control group study the learning conducted with a device developed by the teachers themselves. Therefore, for the students ability the measuring instrument is already included in the device developed. Thus when the evaluation was conducted using the same instrument that is an instrument that has been validated by experts. While the study subjects were students of grade I in elementary school Serayu consisting of 56 learners.

III. RESEARCH RESULTS

The control class taken were as many as 28 learners. Based on the results of the control class pretest it is known that the highest value of cognitive learning outcomes of students obtained with a grade of 83.33 and the lowest grade obtained with a grade of 70.00. Subjects classified as all completed due to meet minimum completeness criteria (Kriteria Ketuntasan Minimal, KKM) that is 6.50. The mean grade of cognitive learning outcomes of students achieved was 76.43 with a standard deviation of 4.15. Based on the results of the posttest the highest grade of control group cognitive learning outcomes obtained with a grade of 86.67 and the lowest grade obtained with a grade of 76.67. Subjects classified as all completed due to meet KKM that is 6.50. The mean grade of cognitive learning outcomes of students achieved was 81.55 with a standard deviation of 3.33.

Meanwhile for the experimental group were also taken as many as 28 students. Based on the data of pretest outcomes it is known that the highest grade of cognitive learning outcomes of learners obtained by the grade of 86.67 and the lowest grade obtained with a grade of 73.33. All students are all completed because it meets the KKM that is 6.50. The mean grade of cognitive learning outcomes of students achieved was 76.07 with a standard deviation of 4.16. Then based on data from posttest it is known that the highest grade of cognitive learning outcomes of learners obtained by the grade of 93.33 and the lowest grade obtained with a grade of 73.33. All students are all completed because it meets the KKM that is 6.50. The mean grade of cognitive learning outcomes of students achieved was 84.76 with a standard deviation of 4.66.

Table 1. Summary of Cognitive Learning Outcomes Grades Data of Students

Details	Cognitive Learning Outcomes of Students					
	Control Class (CC)			Experiment Class (EC)		
	<i>Pretest</i> <i>t</i>	<i>Posttest</i> <i>t</i>	<i>Gain</i> <i>standard</i>	<i>Pretest</i> <i>t</i>	<i>Posttest</i> <i>t</i>	<i>Gain</i> <i>standard</i>
Highest Grade	83,33	86,67	5,11	86,67	93,33	8,69
Lowest Grade	70,00	76,67	6,67	73,33	76,67	3,34
Mean	76,43	81,55	5,12	76,07	84,76	8,69
Deviation Standard	4,15	1,88	3,21	4,16	4,66	3,55

Based on Table 1 it can be seen that the highest grade of pretest acquired by learners in CC that is 83.33, while the lowest grade was obtained by students in CC that is 70.00. The mean grade of pretest at CC is 76.43. The highest grade of posttest at CC is 86.67, the lowest grade is 76.67, with a mean grade of posttest at CC is 84.76. The highest grade of pretest at EC is 86.67, the lowest grade is 73.33, with a mean of pretest is 76.07 on EC. The highest grade of posttest is 93.33 at EC the lowest grade is 76.67, with a mean posttest is 84.76. At CC the increase obtained in cognitive learning outcomes is 3.20 and at EC the increase obtained is 3.55. From Table 49 illustrated that the mean grade of the learners who use the sociocultural-based thematic-integrative learning device is higher than the students who use learning devices used by teachers in the control class.

A study said to be good if the students were able to master the competencies set out in the learning objectives. One indicator of mastery of competencies is how much the grade of learners compared with the KKM. Each educational unit can determine how big the KKM is by considering various aspects. Serayu Elementary School State, individually the learners said to have completed their study if the obtained value of the cognitive learning is at least 70, while classically the learning is said to be completed when 80% of learners completed individually. The grade of cognitive learning outcomes in limited trial aims to determine the completeness of learners.

To determine whether there is difference in the achievement of cognitive learning outcomes of students in both classes, t-test analysis was used. Because the data were normally distributed and homogeneous then the t-test (independent sample t-test) was used. In this study, the calculation of independent sample t-test use the assistance program SPSS version 16.0. The results of the analysis of differences in the results are described as follows.

Hypothesis testing is done on the cognitive learning outcome of cognitive learners (standard gain). Standard Gain data guarantee that the data obtained is purely derived from the treatment and does not come from treatment prior to the experiment. Briefly, the results of calculations independent sample t-test for the two groups in terms of cognitive learning outcome of students can be summarized in the following table.

Table 2. The Test of the Differences of Cognitive Learning Outcomes Improvement of Students

Class	Mean grades	N	Db	T _{count}	t _{table}	p	Ket
EC	86,90	28	54	3,95	2,670	0,0001	H ₀ rejected
CC	51,19	28					

Table 2 above shows that t_{count} is 3.95 with a significance level of 0.0001. Because $t_{count} > t_{table}$ or significance value less than 0.05, then H_0 is rejected. Based on these data it can be concluded that there is a significant difference in the improvement of cognitive achievement among students who follow learning by using socio-cultural-based learning devices than students who take the conventional learning. Based on Table 2, the initial ability of students in both classes are relatively similar. The mean grade of pretest EC is only slightly higher than with CC. But if we look more closely, the average pretest grade higher on EC because there are some students who have 'accidental' grades that is high enough on the tests. It is said to be 'accidental' because the test scores in post-test has decreased somewhat drastic. If viewed from the post-test, EC has an mean grade higher than with CC. That's because the EC grade on the test increased significantly compared to CC, although the grade of CC on the test slightly higher than CC. With reference to the pre-test and post-test, it appears that the average standard gain of EC is higher than CC. The use of standard gain in this analysis is to illustrate that the increase in the cognitive learning outcomes grades caused by the treatment during this research process and not by the previous treatment.

Based on the t-test results, statistically seen that there are differences in the standard gain cognitive learning outcomes between EC and CC. That's because in the EC, the whole process of learning is based on the development of cognitive, affective, and psychomotor. This was reinforced by the use of thematic-integrative learning that integrates a variety of skills from different fields of study in a single theme. At CC, the main base is on cognitive aspects of learning with C1-dimensional level (knowledge), C2 (understanding), C3 (application), and very little portion C4 (analysis), C5 (synthesis) and C6 (creation). Teachers always give exercises related to the concept being studied. It made the students are trained to solve understanding problems so it is not surprising if the mean grade of CC learners was also good. In addition, teachers are also always working to make learning contextual with the real conditions, but these efforts have not run optimally. Learners do not participate more actively so that the psychomotor aspects of the mean grade of CC is not so good. It can be concluded that in total, the grade of cognitive learning outcomes of EC learners is better than CC learners.

IV. DISCUSSION

Learning device has a good effectiveness on the syllabus, lesson plans (Rencana Program Pembelajaran, RPP), learning media, and tests of cognitive learning outcomes. The syllabus was developed to make teachers more quickly in setting up a syllabus with more comprehensive material. RPP developed to make teachers more quickly in preparing learning activities in a coherent and comprehensive manner. Learning media developed to make teachers become easier to demonstrate learning materials. Test of cognitive learning outcomes was developed to make the teacher becomes easier and faster in evaluating the ability of learners.

The average achievement of the cognitive learning has increased in terms of the scores of pre-test and post-test. At the same time EC experienced an increase higher than CC. Based on t-test results, statistically seen that there are differences in the standard gain cognitive learning outcomes between EC and CC. That's because in the EC, the whole process of learning is based on the development of cognitive, affective, and psychomotor. This was reinforced by the use of thematic-integrative learning that integrates a variety of skills from various fields of study within

the sub-themes. Teachers always give the exercises related to the concept being studied. It made the students are trained to solve understanding problems it is not surprising if the mean grade of CC learners was also good. In addition, teachers are also always working to make learning contextual with the real conditions, but these efforts have not run optimally. It can be concluded that in total, the grade of cognitive learning outcomes of EC learners is better than CC learners. Overall as a learning device unit has a high effectiveness that is able to make all students (100%) achieved complete learn.

From observations on the limited testing and field testing that the activity of students looked excited, so excited, learners are also enthusiast, especially when conducting experiments outside the classroom, discuss, discover or find information in their neighborhood. The increase of cognitive learning outcomes obtained and positive response as well as the learning process are fun for learners due to several things, among others: 1) the learning device used by learners is different from the previous learning; 2) Contextual learning process, not only in the classroom but also in the environment of learners; 3) learning activities engage learners actively; 4) learning resources and learning media used are very close with the live of learners; and 5) learners do the learning process not only with teachers, but also with peers, parents, adults, environment and culture. Learning device used is said to be different because the device is sociocultural-based. The learning process is not only monotonous in the classroom but also outside the classroom and in the community environment of learners.

This is consistent with Vygotsky's theory that emphasizes the sociocultural nature of learning. Two main implications of Vygotsky's theory of learning. First, the class structures wished to be organized as cooperative learning among students, so that students can interact around difficult tasks and each raises effective problem-solving strategies in each of their zone of proximal development. Cooperative learning is realized through group learning activities that have been implemented consisting of 5 to 6 students. Second, the Vygotsky approach in teaching emphasized the scaffolding so that students are increasingly has a responsibility toward their own learning. For example, in the reciprocal teaching, teachers lead small groups of students to ask questions about the reading text they have read and gradually transfer responsibility to lead the discussion to the learners.

According to Vygotsky, the social interaction in cognitive development is important. Human cognitive development is closely related to language development. Because language is a strength for human mental development. Cognitive development depends on how much learners actively manipulate and interact with their environment. This indicates that the environment in which students learn largely determines the cognitive development of learners.

Vygotsky also noted that a person's interactions with the environment can help learning. The experiences that brought a person (in this case the students) to a learning situation can greatly affect the cognitive learning. When learners alongside their peers work together on tasks, their social interactions together can act as a teaching function. Constructivism environment is best designed for meaningful learning and structured in-depth, not just for surface understanding.

Vygotsky believed that people and their cultural environment plays a role in the system of social interaction. Through communication and action, people who are in the neighborhood teach their children the tools (for example, the language of symbols, signs) they need to acquire competencies (Schunk, 2012: 581). Social interaction with teachers, parents and peers who are more experienced gives a significant contribution to the intellectual development of children. This was confirmed by Jackman (2012: 10) as cited below.

"Much of what a child learns comes from the culture around him. In addition, interactions with teachers, parents, and more experienced peers contribute significantly to a child's intellectual development."

According to Vygotsky children construct knowledge through social interaction. Even though this was not in line with Piaget that stating that children construct knowledge by doing transformation, organizing and reorganizing the prior knowledge (Santrock, 2011: 66). Vygotsky also emphasized the role of language in the thought process. He considered that the importance of language in everyday life of children. They need a language to talk with others, listen to others, reading and writing. Language allows them to describe past events in detail and to plan for the future. Language allows them to take information from one generation to the next and creating a rich cultural heritage. Similarly through the language cultural values that exist in the environment

of learners will not lost and will not recognized by the learner as a future successor. As noted by Warsono and Hariyanto (2013: 59) that the assumptions key of the theory of Vygotsky is "What the child can do in cooperation today he can do alone tomorrow." From all aspects said before supported by the sociocultural-based thematic-integrative device developed in this study.

In addition to the learning process, which is based on constructivism, on thematic-integrative learning authentic assessment was also conducted. Authentic assessment form will require learners to make writing of the submission of the mind, discuss what they have learned and why this knowledge is useful in the world or demonstrate and apply the skills that have been acquired. Authentic assessment have also been developed, contained in the RPP developed in this study.

The uniformity between theory and applications that occur in the field during the research process becomes the basis that the cognitive development of students has increased. One form of cognitive learning outcomes can be seen from the test of cognitive learning outcomes of students who are experienced individual completeness. In addition to some of the matters described above, the success or completeness and improvement of cognitive learning outcomes of students achieved also shows that the process of learning use a scientific approach, as designed by researchers has a significant influence. It can also be seen from the results of authentic assessment obtained by learners is very good, good performance appraisal, project assessment, portfolio assessment and written assessment. From these things it can be concluded that learning using the sociocultural-based thematic-integrative learning device can improve cognitive learning outcomes of learners.

In the process of thematic-integrative learning using learning tools that is the result of the development of field tests also found that the ability of learners to apply their learning in everyday life is increasing. This is evidenced by the increase in cognitive learning outcomes that is seen at pre-test and post-test. Learners are able to apply the knowledge, skills, and behavior in solving everyday problems, as well as the learners are also more susceptible to observe phenomena that exist in the surrounding environment.

Thus, by the end of the study it can be said that the learning tools that is the result of the development is a product that has been viable for use in thematic-integrative learning in the field. Another characteristic of the sociocultural-based learning with a sub theme of heroism result of this development are some of the advantages it has. These advantages include: (1) the learning device containing devices, all of which contain elements of sociocultural, so that learning can not be separated from the cultural life of the learners, (2) learning will be more effective, efficient and enjoyable as all the main learning required is already exist in this learning tool.

V. CONCLUSIONS AND SUGGESTIONS

Based on the results and discussion of this study results, it can be overall concluded as a unitary learning device on the theme of my experience are highly effective for improving the academic performance of students. Learning device that can improve cognitive learning outcomes in learning, then the utilization required a collaboration between teachers, school leaders and national education department party to provide the necessary facilities. this is necessary because in the implementation the learning device utilization requires facilities and additional costs when compared to conventional learning.

REFERENCES

- [1] Jackman, H.L. (2012). *Early educational curriculum: a child connection to the world*. Bemidji: Cengage Learning
- [2] Kyriacou, C. (2009). *Effective teaching in school (theory and practice)*. London: Nelson Thornes.
- [3] Nur Wangid, M., Mustadi, A., dan Astuti, P.A. (2014). *Pengembangan perangkat pembelajaran tematik-integratif sekolah dasar berbasis sosiokultural di Daerah Intimewa Yogyakarta*. PPs UNY. Research Report. Unpublished.
- [4] Santrock, J.W. (2011). *Educational psychology*. New York: McGraw Hill.
- [5] Suhadi. (2007). *Belajar aktif dan terpadu*. Surabaya: Duta Graha Pustaka.
- [6] Sutarno. (2007). *Pendidikan multikultural*. Direktorat Jenderal Pendidikan Tinggi: Departemen Pendidikan Nasional.
- [7] Schunk, H.D. (2012). *Learning theories an educational perspective*. Boston: Pearson.
- [8] Warsono dan Hariyanto (2013). *Pembelajaran aktif teori dan asesmen*. Bandung: PT Remaja Rosdakarya

THE INFLUENCE OF TEACHER PROFESSIONAL COMPETENCE AND INTERPERSONAL INTELLIGENCE TOWARDS MOTIVATION AND PHYSICS LEARNING RESULT OF STUDENT AT XI MIA GRADE SMA NEGERI 1 PANGKAJENE

Murniaty M¹, Kaharuddin Arafah², Subaer³

¹Student at Phisycs Education Post Graduate Program, State University of Makassar

²Lecturer at Phisycs Education Post Graduate Program, State University of Makassar
murni.mursalin@yahoo.com

Abstract - This research is ex-post-facto research which aims to determine the influence of: (1) the teacher professional competence towards physics learning result; (2) interpersonal intelligence towards physics learning result; (3) the teacher professional competence towards learning motivation; (4) interpersonal intelligence towards learning motivation; and (5) learning motivation towards physics learning result. The populations in this study were all students of XI MIA Grade SMA Negeri 1 Pangkajene as many as 173 students. The sample taken by using Slovin technique with 122 students. The process of collecting data using questionnaires and test of physics learning result that have been tested empirically. Data of the research result were analyzed by using analysis method of *Structural Equation Modeling* (SEM) with technique of *Analysis of Moment Structures* (AMOS). The analysis procedure is performed with descriptive analysis and inferential analysis, factor analysis and verification of structural model AMOS. The result of research showed that the structural equation model that describes the influence of teacher professional competence, interpersonal intelligence, learning motivation, and physics learning result can be accepted. Through the model can be concluded that: (1) the teacher competence professional doesn't have positive direct influence and significant toward physics learning result; (2) interpersonal intelligence doesn't have positive direct influence and significant towards physics learning result; (3) the teacher professional competence has positive direct influence and significant towards learning motivation; (4) interpersonal intelligence has positive direct influence and significant towards learning motivation; and (5) learning motivation has positive direct influence and significant towards physics learning result.

Keywords: *ex-post-facto, teacher competence professional, interpersonal intelligence, learning motivation, physics learning result*

I. INTRODUCTION

The progress of a nation is determined by the quality of human resources, and quality of human resources depends on the quality of education. Education is one of the main pillars of building a nation towards modern civilization, and handles a very important and strategic for human life, because human beings through education will be equipped with a range of capabilities to meet the challenges and changes through a learning process.

In this case the teacher is the most decisive component in the education system as a whole that should be a major concern. The teachers handle a major role in the development of education, particularly organized formally in school. The teacher is a component that most influence on the creation process and the quality of educational outcomes. Reference [1] states that the learning process and learning outcomes of students is not only determined by the school, patterns, structures, and the curriculum, but is largely determined by the competence of

teachers who teach and guide students. Competence of teachers is one of the factors that affect the achievement of learning objectives and education in schools. One of teacher competence will be examined in this study, namely professional competence. Professional teachers are teachers who have the ability and expertise in the field of teacher training so that they can perform their duties and functions to provide knowledge and character to the students to the maximum.

Keep in mind other than IQ that developed at this time, there are many types of intelligence that go with it, such as intelligence concept coined by Howard Gardner. Of the many intelligence delivered by Gardner (2003) in the theory of Multiple Intelligence, he suggests the importance of interpersonal intelligence. Interpersonal intelligence is the intelligence that involves the skills to work with others and communicate with both verbal and non-verbal [2]. In addition to high Intelligence Quotient, interpersonal intelligence is also required in the process of learning physics are not just studying the concept and doing experiments. Not quite up there, not all of the material in physics can be solved by the individual personally, sometimes students need a teacher or a friend to explain associated with such materials. By understanding the interpersonal intelligence will assist students in the process of studying physics.

The success of students in developing interpersonal intelligence is also influenced by the presence of psychological factors that exist in self-students. Psychological factors will lead to the desire, drive and passion for learning, or move the students to study harder. The psychological factors include the motivation to learn, motivation to learn also give good directions to the learning activities undertaken by students, so that the desired learning result can be achieved.

Thus the rationale encourage researchers to conduct research on the influence of teachers' professional competence and interpersonal intelligence towards motivation and physics learning result of students at XI MIA Grade SMA Negeri 1 Pangkajene. The problems in this research are: (1) Does the teacher professional competence have positive direct influence towards physics learning result? (2) Does interpersonal intelligence have positive direct influence towards physics learning result? (3) Does the teacher professional competence have positive direct influence towards learning motivation? (4) Does interpersonal intelligence have positive direct influence towards learning motivation? (5) Does the learning motivation have positive direct influence towards physics learning result? The purpose of this study was to answer the formulation of the problem, namely (1) to determine the direct influence of teacher professional competence towards physics learning result, (2) to determine the direct influence of interpersonal intelligence towards physics learning result, (3) to determine the direct influence of teacher professional competence towards learning motivation, (4) to determine the direct influence of interpersonal intelligence towards learning motivation, and (5) to determine the direct influence of learning motivation towards physics learning result. The results of this study are expected to provide good benefits for students, teachers, schools and other researchers. For students, this study can provide input to further improve learning result. For teachers, this research as information can provide a positive discourse. For schools, this research can be used as information material in improving the teaching competence of teachers. As for other researchers, this study can be used as reference for further research, especially research that focuses on internal factors on students to subjects Physics.

II. RESEARCH METHOD

The type of research is the study "ex post facto", which is causality and correlation. This study attempted to investigate the direct influence of the independent variables namely the teacher professional competence and interpersonal intelligence on physics learning result as dependent variables, both directly and through the learning motivation as an intervening variable. Designs of linkages between these variables are described as follows.

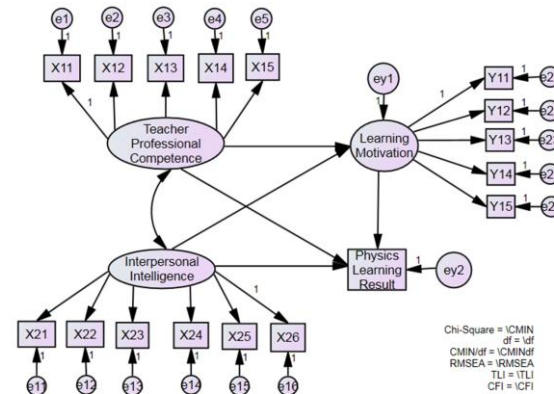


Figure 1. Structural Model of Functional Interwoven Proposed Among Variables Research

The populations in this study were all students of class XI MIA SMA Negeri 1 Pangkajene academic year 2015/2016, Pangkep Regency, and South Sulawesi. Distribution of students in each class is shown in table 1 below.

Table 1. Distribution Of Student Class XI MIA

Name of Class	Number of Students
XI MIA 1	33
XI MIA 2	35
XI MIA 3	35
XI MIA 4	35
XI MIA 5	35
Total	173

The technique used to determine the sample size is by slovin technique. As for the distribution of the sample is more presented in table 2 below.

Table 2. Distribution of Research Sample

Name of Class	Number of Students	Sample
XI MIA 1	33	-
XI MIA 2	35	32
XI MIA 3	35	33
XI MIA 4	35	28
XI MIA 5	35	29
Total	173	122

The instrument used in this study consisted of questionnaire of the teacher professional competence, interpersonal intelligence questionnaire, questionnaire learning motivation, and test of physics learning result.

Before the instrument is ready for use, it must first be validated instrument that is validation of content and empirical validity. Contents validation test conducted on measuring instruments. Analysis of the contents of the instrument validation is done in this research that uses models Gregory in the form of a model agreement among experts. Empirical validation test performed on the test results the instrument consisting of item validity test and reliability test. Types and data collection techniques used in this study can be seen in the following table.

Table 3. Types and Data Collection Techniques

Data	Type of Data	Data Collection Techniques	Data Sources	Instrument
Teacher Professional Competence	Interval	Questionnaire	Students	Questionnaire of Teacher Professional Competence
Interpersonal Intelligence	Interval	Questionnaire	Students	Questionnaire of Interpersonal Intelligence
Learning Motivation	Interval	Questionnaire	Students	Questionnaire of Learning Motivation
Physics Learning Result	Interval	Questionnaire	Students	Questionnaire of Physics Learning Result

Analysis of the data used in this study consisted of a statistical analysis of descriptive and inferential analysis to test the hypothesis.

A. Analysis of Descriptive Statistic

A function of descriptive analysis is to provide an overview of the data obtained, such as: the number, maximum, minimum, mean, mode, median, standard deviation and variance.

B. Normality Test

Normality test is used to determine whether the data sample studied came from populations with normal distribution or not. Values were considered in determining test multivariate normality is critical ratio by using AMOS 22.0.

C. Linearity Test

Linearity test is performed to determine whether there is a linear relationship between the independent variables with the dependent variable using SPSS 22.0. The variable is said to be linear with other variables when sig.Linearity > 0.05.

D. Multicollinearity Test

Multicollinearity test is performed to determine whether there is a significant relationship between independent variables in a multiple linear regression model. Multicollinearity test was performed using statistical program SPSS 22.0 to see the value of Tolerance and Variance Inflation Factor (VIF) in the regression model.

E. Factor Analysis

Factor analysis was performed using AMOS 22.0 to be able to test the influence of the indicator with latent variables, a model must be eligible Goodness of Fit, which is an index that is used as a reference model is said to be acceptable fit. The index used is the Chi-square, CMIN / df, TLI, CFI and RMSEA.

III. RESULT AND DISCUSSION

A. Analysis Result of Research Data

1. Descriptive Statistic

Table 4. Summary of Descriptive Analysis Result

Statistic	Teacher Professional Competence	Interpersonal Intelligence	Learning Motivation	Physics Learning Result
Number of respondent	122	122	122	122
Number of item	30	32	30	29
Mean	112.50	121.22	110.95	15.69
Std. Error of Mean	.985	1.098	1.399	.490
Median	113.00	121.00	111.00	16.00
Mode	111.00 ^a	128.00	105.00 ^a	9.00 ^a
Deviation Standard	10.88	12.13	15.46	5.41
Variance	118.417	147.154	239.022	29.337
Skewness	-.765	-.268	-.333	-.116
Kurtosis	2.348	.240	.127	-.987
Range	70.00	67.00	81.00	22.00
Minimum	64.00	84.00	63.00	4.00
Maximum	134.00	151.00	144.00	26.00
a. Multiple modes exist. The smallest value is shown				

Research result data of variable the teacher professional competence then presented the list of frequency distribution as shown in table 5 below.

Table 5. Distribution of Frequency, Percentage, and Categories for Teacher Professional Competence

Interval Score	Category	Frequency	Percentage (%)
30.00 – 54.40	Very Low	0	0
54.50 – 78.40	Low	1	0.82
78.50 – 102.40	Moderate	17	13.93
102.50 – 126.40	High	93	76.23
126.50 – 150.00	Very High	11	9.02
Total		122	100

Research result data of variable the interpersonal intelligence then presented the list of frequency distribution as shown in table 6 below.

Table 6. Distribution of Frequency, Percentage, and Categories for Interpersonal Intelligence

Interval Score	Category	Frequency	Percentage (%)
32.00 – 58.00	Very Low	0	0
58.10 – 83.60	Low	0	0
83.70 – 109.20	Moderate	20	16.40
109.30 – 134.80	High	86	70.49
134.90 – 160.00	Very High	16	13.11
Total		122	100

Research result data of variable the learning motivation then presented the list of frequency distribution as shown in table 7 below.

Table 7. Distribution of Frequency, Percentage, and Categories for Learning Motivation

Interval Score	Category	Frequency	Percentage (%)
30.00 – 54.40	Very Low	0	0
54.50 – 78.40	Low	3	2.46
78.50 – 102.40	Moderate	30	24.60
102.50 – 126.40	High	73	59.83
126.50 – 150.00	Very High	16	13.11
Total		122	100

Research result data of variable physics learning result then presented the list of frequency distribution as shown in table 8 below.

Table 8. Distribution of Frequency, Percentage, and Categories for Learning Motivation

Interval Score	Category	Frequency	Percentage (%)
0.00 – 6.20	Very Low	2	1.64
6.30 – 12.00	Low	33	27.05
12.10 – 17.80	Moderate	38	31.14
17.90 – 23.60	High	41	33.61
23.70 – 29.00	Very High	8	6.56
Total		122	100

2. Analysis Prerequisites Test

a. Data Normality Test

By using a significance level of 0.01, the data is said to be normally distributed if the critical ratio (cr) of kurtosis is between ± 2.58 . Based on the results of the output data normality test on Assessment of normality, normality test results obtained by the value of the multivariate cr kurtosis $3.497 > 2.58$ which means multivariate distribution is not normal.

To make the normal distribution of data, then the next step is the possibility of detecting the data included in the category of outlier by looking at the table Mahalanobis Distance. Thus, most data outliers and should be excluded from the data analysis is the number of respondents are 53 and 43.

After removing outlier values obtained cr multivariate kurtosis 1.889. It shows the data is normally distributed, either individually or multivariate and deserves to be analyzed further.

b. Linearity Test

Based on the results of linearity test the influences of X_1 with Y_2 have sig Linearity 0.000. For the influence of X_2 with Y_2 have sig Linearity 0.000. As for the influence of Y_1 with Y_2 have a sig Linearity 0.000. This means that the value of sig Linearity less than 0.05 ($0.000 < 0.05$). So it can be concluded that the influence of variable professional competence of teachers with physics learning result, the influence interpersonal intelligence with physics learning result and learning motivation with physics learning result are linear and are eligible for further analysis.

c. Multicollinearity Test

Table 9. The Result Of Multicollinearity Test

Model	Collinearity Statistics	
	Tolerance	VIF
Teacher Professional Competence (X_1)	0.777	1.287
Interpersonal Intelligence (X_2)	0.603	1.659
Learning Motivation (Y_1)	0.551	1.816

Dependent Variable: Physics Learning Result (Y_2)

3. Factor Analysis

a. Factor Analysis of Latent Variables

Test the fit between the theoretical models to empirical data can be seen at the level of Goodness of Fit Statistics.

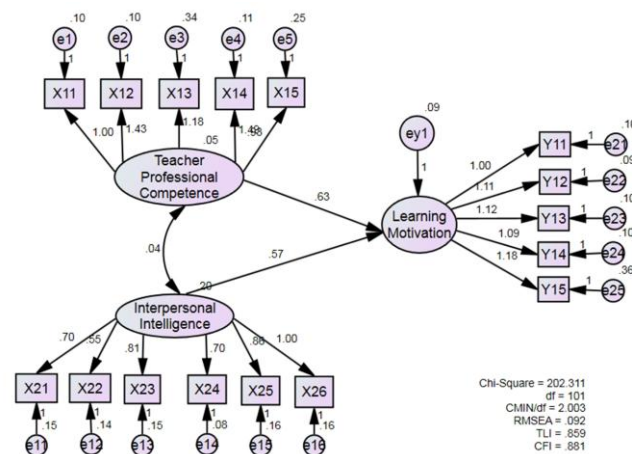


Figure 2. Initial Factor Model of Latent Variables

The result of the factor analysis beginning shows that there are indices that do not meet the cut off value. Modifications made some errors influence of variables that have a value large change Chi-square. The modification result then re-analyzed by the results in Figure3.

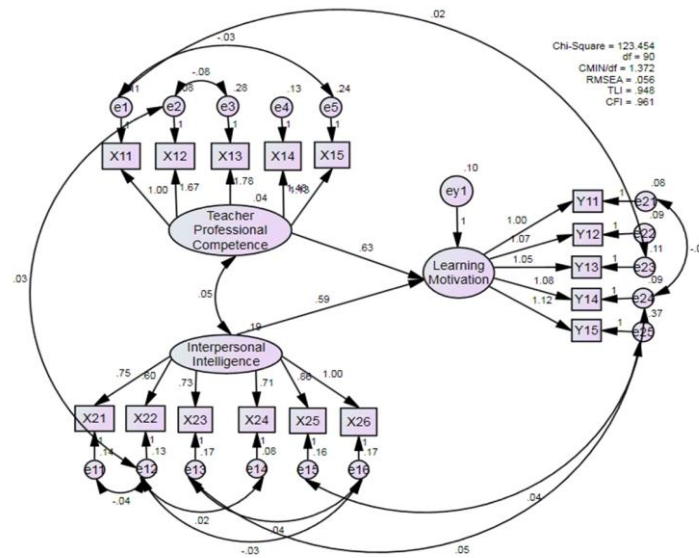


Figure 3. Final Factor Model of Latent Variables

The final result can be seen that all the indexes have met the criteria so that these models can be received and analyzed further. Standardized regression weights can also show the influence of latent variables with the indicators.

Table 10. Standardized Regression Weights Of Latent Variables

		Estimate
X ₁₁	<--- Teacher Professional Competence	0.853
X ₁₂	<--- Teacher Professional Competence	0.858
X ₁₃	<--- Teacher Professional Competence	0.834
X ₁₄	<--- Teacher Professional Competence	0.859
X ₁₅	<--- Teacher Professional Competence	0.655
X ₂₆	<--- Interpersonal Intelligence	0.733
X ₂₅	<--- Interpersonal Intelligence	0.688
X ₂₄	<--- Interpersonal Intelligence	0.735
X ₂₃	<--- Interpersonal Intelligence	0.616
X ₂₂	<--- Interpersonal Intelligence	0.586
X ₂₁	<--- Interpersonal Intelligence	0.658
Y ₁₁	<--- Learning Motivation	0.522
Y ₁₂	<--- Learning Motivation	0.764
Y ₁₃	<--- Learning Motivation	0.566
Y ₁₄	<--- Learning Motivation	0.639
Y ₁₅	<--- Learning Motivation	0.428

1) Model Verification and Final Model Development

Verified the theoretical model developed based on empirical data. An analysis of the Figure 3 is the structural equation model of initial stages.

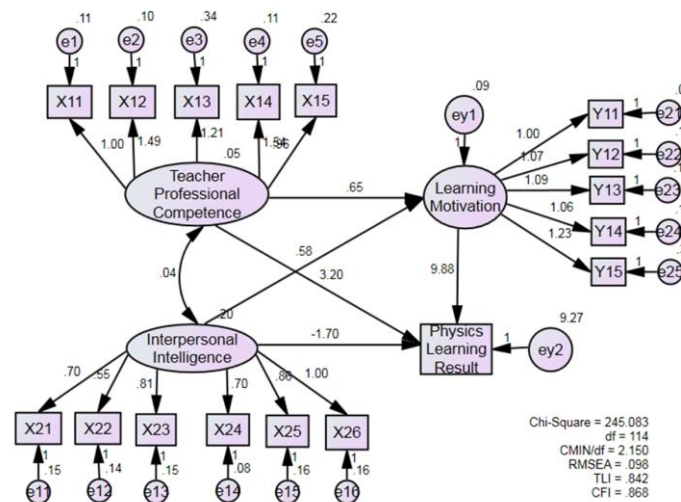


Figure 4. Structural Equation Model of Initial Stages

The results of the analysis of the initial stages in figure 4 shows that there are indices that do not meet the cut off value. Modifications made some errors influence of variables that have a value large change Chi-square. The modification result then re-analyzed by the results in Figure 5.

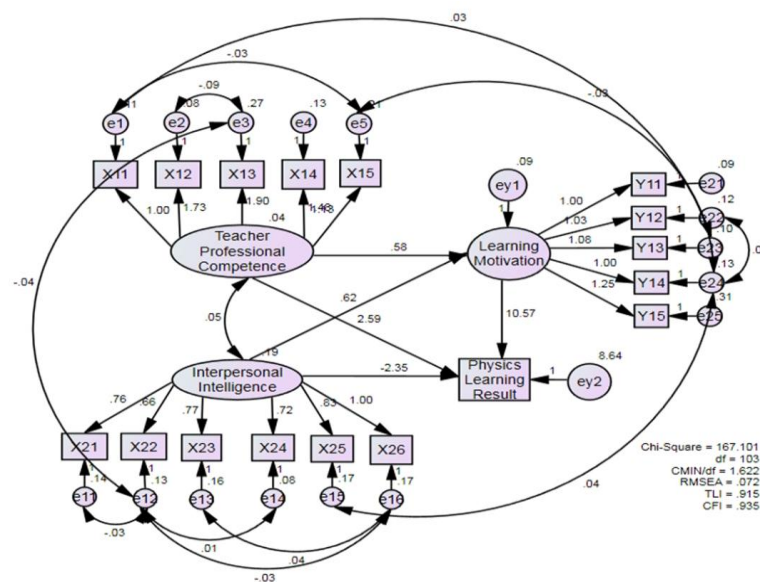


Figure 5. Structural Equation Model of Final Stages

In the final result can be seen that all the indexes have met the criteria so that these models can be received and analyzed further. Parameter of regression weighted shown in the table 10 below.

Table 11. Regression Weight of Final Model

			Estimate	S.E.	C.R.	P
Y ₁	<---	X ₁	0.578	0.241	2.401	0.016
Y ₁	<---	X ₂	0.621	0.125	4.947	0.000
Y ₂	<---	Y ₁	10.565	1.394	7.579	0.000
Y ₂	<---	X ₂	-2.341	1.354	-1.729	0.084
Y ₂	<---	X ₁	2.608	2.264	1.152	0.249

Structural model is obtained based on the index overall fit can be seen in table 10. Mathematically structural equation model can be written:

$$\hat{Y} = 15.683 + 2.608X_1 - 2.341X_2 + 10.565Y_1 \quad R^2 = 70\%$$

B. Discussion

1. The Direct Influence of Teacher Professional Competence towards Physics Learning Result.

For the first hypothesis testing showed that the influence variables of teacher professional competence towards physics learning result described in the regression weights of final model with the estimate results $\beta_{x_1y_2} = 2.608$ with p value = $0.249 > 0.05$. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that the teacher professional competence does not have positive direct influence and significant towards physics learning result.

2. The Direct Influence of Interpersonal Intelligence towards Physics Learning Result

For the second hypothesis testing showed that the influence variables of interpersonal intelligence towards physics learning result described in the regression weights of final model with the estimate results $\beta_{x_2y_2} = -2.341$ with p value = $0.084 > 0.05$. This means that H_0 accepted and H_1 rejected at the significance level 0.05. This result indicates that interpersonal intelligence does not have positive direct influence and significant towards physics learning result.

3. The Direct Influence of Teacher Professional Competence towards Learning Motivation

For the third hypothesis testing showed that the influence variables of teacher professional competence towards learning motivation described in the regression weights of final model with the estimate results $\beta_{x_1y_1} = 0.578$ with p value = $0.016 < 0.05$. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that teacher professional competence has positive direct influence and significant towards learning motivation.

4. The Direct Influence of Interpersonal Intelligence towards Learning Motivation

For the fourth hypothesis testing showed that the influence variables of interpersonal intelligence towards learning motivation described in the regression weights of final model with the estimate results $\beta_{x_2y_1} = 0.621$ with p value = $0.000 < 0.05$. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that interpersonal intelligence has positive direct influence and significant towards learning motivation.

5. The Direct Influence of Learning Motivation towards Physics Learning Result

For the fifth hypothesis testing showed that the influence variables of learning motivation towards physics learning result described in the regression weights of final model with the estimate results $\beta_{y_1y_2} = 10.566$ with p value = $0.000 < 0.05$. This means that H_0 rejected and H_1 accepted at the significance level 0.05. This result indicates that learning motivation has positive direct influence and significant towards physics learning result.

IV. CONCLUSION AND SUGGESTION

The results showed that: 1) the teacher professional competence does not have positive direct influence and significant towards physics learning result; 2) interpersonal intelligence does not have positive direct influence and significant towards physics learning result; 3) the teacher professional competence has positive direct and significant towards learning motivation; 4) interpersonal intelligence has positive direct influence and significant towards learning motivation, and 5) the learning motivation has positive direct influence and significant towards physics learning result.

Based on the research results obtained, it is advisable that matters; 1) For teachers to better understand the psychological factors that can affect the results of students; 2) for students to further develop the factors that was in him to get the optimal learning results; 3) For researchers interested in developing further this research, is expected to examine the limitations in this study, so further research can enhance this research.

REFERENCES

- [1] Hamalik, Oemar. 2004. *Pendidikan Guru Berdasarkan Pendekatan Kompetensi*. Jakarta: Bumi Aksara.
- [2] Alder, Harry. 2001. *Boost Your Intelligence: Pacu EQ dan IQ Anda*, Terj. Christina Priarningsih. Jakarta: Erlangga

EVALUTION STUDY TO CAREER GUIDANCE SERVICE-PROGRAM OF VOCATIONAL HIGH SCHOOLS IN BANJARMASIN

Nina Permatasari¹, Djaali², Ma'ruf Akbar³

Counselling and Guidance Department, FKIP Of Lambung Mangkurat University
Research and Education Evaluation Department, Universitas Negeri Jakarta
E-mail: ninapermatasari1980@gmail.com

Abstract - The aim of the study was to evaluate career guidance service-program in vocational high schools in banjarmasin. This evaluation study used *Stufflebeam-CIPPO* (*Context, input, process, product, outcome/sumatif*) model and summative evaluation used *Scriven-summative evaluation*. Repondents in this study were Headmasters, counselors, teachers, homeroom teachers, students, business and industrial partnerships that have a collaboration relationship with vocational schools. Data collection techniques used by observation, questioner, interview and documents study. Data analysis used qualitative and quantitave descriptive technique. Study results showed: evaluation context was in 'poor' category (59,62%); input evaluation from teachers questioner was in 'very good' category (87,85%); input evaluation from students questioner was in 'good' category (75,64%); process evaluation in implementation program was in 'good' category (79,74%); program-delivery aspect was in 'good' category; media-used aspect was in 'good' category; relationship to industry and school monitoring was in 'good' category (84,94%); product evaluation was in 'good' category (72,05%); maturity-career of students was in 'good' category (69,25%). Outcome/summative from qualitative data showed was in 'good' category.

Keywords : *program evaluation, career counselling program, CIPPO*

I. INTRODUCTION

Career guidance service-program is a guidance process, service and approachment to student so they will know and understand their self better, concerding about work, making a future plan base on what they expect, taking and trusting theirbest desicion base on their condition if we relate it to requirement of job and demand of job or career they choose¹.

Career guidance service-program of vacational school in Banjarmasin doesn't work effectively because the implementation is not optimal yet. It means the service doesn't follow the formal rule of the implementation of career guidance service-program in shool. The realization of this program in school is only get small part of school program compare to other three guidance program (self guidance program, social guidance program and, study guidance program). It is a must for Counselor in every school to evaluate the career guidance service program because by evaluating it, Counselor will get some informations or explaintaion about the lack of program or the program ineffectiveness during the implementation. So, Counselor will know which component should be fixed, should be left, even should be removes if that component harm the implementation of the whole program.

Generally, this research aim to get information and explanation about the implementation of career guidance service program of vacational school in Banjarmasin, specifically this research aim to : 1) Describe the need and aim of program in context step, 2) Find of the readiness related to schedule and activity of program, student recruitment, social economic of student family, education/school stakeholder, facilities and infrastucture and also program budgeting in input step, 3) Find out the realization of program by observing the process, cooperation, supervision and monitoring in process step, 4) Find out the target reached result and student understanding

to choose their career in product step, 5) Find out the impact of program toward student target graduation, ability of alumnus to choose their career or job and develop and improve their self in entrepreneurship in sumative and outcome step.

A. Materials

Review of literature are the definition of evaluation, evaluation program, CIPPO Sumative model, and career guidance service program. Tayler in Brinkerhoff² explains that evaluation is a process to compare between aim that already set before or the implementation of program and aim that reach after the process or the implementation of program.

Evaluation model are the combination of CIPPO evaluation method developed by Scriven, through five step which are context evaluation, input evaluation, process, produk dan Outcome/Sumatif.

Career guidance service program is a program to prepare student to facing work world, choosing job and giving self investment so student ready to get any position and also in self adapting to job demand.

Generally the purpose of assessment in career program guidance of vocational school includes: Self and surrounding understanding, self understanding includes ability/skill identification and their life values to develop their self. Surrounding understanding includes the knowledge about physical, social, culture, information, surrounding (information, education, career and individual social).

Capable to deciding, planning their life in education, career and individual social. Developing their capability and willingness maximally. Solving individual problem wisely. Understand and guiding their self when acting and having attitude base on demand and surrounding.

II. RESEARCH METHODOLOGY

Evaluation method used in this research is combination of CIPPO evaluation method develop by Stufflebenn and Sumatif evaluation (result evaluation) develop by Scriven.

Data collection techniques used questionnaire, interview guidance, and documentation note, career ripeness career test(inventory), observation guidance.

Data analysis used Miles and Huberman model, data analsing consist by three process which are: data reduction, data display and conclusion drawing/verification⁵. Meanwhile to check the validity use triangulation technique, it means researcher use some different technique in collecting data from the sae resource, in other hand triangulation resource means that researcher use the same technique in different resource

III. RESULT AND DISCUSSION

A. Result

Result of career guidance service program evaluation are :

1. Context Components

Base on overall context evaluation, objective standard criteria toward context component display on tabel below.

Table 1. Context Assessment

Comp.	Aspects	Score %	Assessment description			Conclusion
			Very good	Good	Poor	
Context	Formal legality School vision and mission Aim of program Test/non test implementation Assessment instrument accuracy Data analysis	59,62			√	Having formal legality Vision and mission is understandable for school stakeholder except student Aim of program is not discussed together Counseling and Guidance Department was not taking part to choose program for new student because it was held online Non testing technique for validity and reliability of instrument can't be consider. Report of assessment analysis result reported periodic in one semester

Base on tabel 1 seen that context evaluation in career guidance service program of vocational school in Banjarmasin generally is 59,62% with the category is poor.

2. Input Components

Input component evaluation result display on tabel below.

Table 2. Input assessment

Comp.	Aspect	Score %	Assessment description			Conclusion
			Very good	Good	Poor	
Input	Planning	78,79		√		Program planning is not good for characteristic, way, kind and stage level can't follow the Vocational school curriculum
	Activity and schedule					Planned every year, semester, onth, week and day that written in RPL
	Student recruitment					There is no spesific requirement
	Social economic data of student family/parents					There is Social economic data of student family/parents
	School counselor					Having couelor teacher with rasio 1: 150
	Facility and infrastucture					Classroom used good for the amount of students Counseling room is representative enough There is practice room base on student program
	Budgeting					There is no special budgeting for career guidance service program. Earn from official resource and handled with transparency There is finance report that reported periodic to the headmaster of school

Based on table 3 it can be seen that the results of the evaluation input in career guidance services program at Vocational High Schools (SMK) in Banjarmasin obtain a percentage value of 78.79% which belong to "good" or "adequate".

3. Process Components

Based on the evaluation of the input, the component can be presented in Table 3 below.

Table 3. Assessment of the process

Comp.	Aspect	Value (%)	Assessment Description			Conclusion
			Very good	Good	Poor	
Process	Execution of the event	70,83		√		Execution: Orientation service, placement and distribution information, content mastery, individual counseling and group guidance.
	The material delivery	65,00		√		The delivery of the material was delivered well based on the service plan.
	The media of the service	60,83			√	The media is lack of creativity and didn't use the available technology
	The techniques of the service	58,33			√	The technique which be used is lecture and textbook. There are no various methods.
	Cooperative Monitoring	72,11		√		The cooperative has been tied with all stakeholders in school.

Based on table 3 it can be seen that the results of the evaluation input in career guidance services program at Vocational High Schools (SMK) in Banjarmasin based on five aspects of the implementation process for the enforceability of RPL activities in accordance with the planned counselor 70.83% with good category, delivery of material 65.00 % (good category), media services that are used 60, 83% with poor category, the service method 58.33% with poor category, and for process monitoring and cooperation with school stakeholders 72.11% with both categories.

4. Product Component

Based on the evaluation of the components of the product can be presented in Table 4 below.

Table 4. The assessment of the product

Comp.	Aspect	Value (%)	Assessment Description			Conclusion
			Very good	Good	Poor	
Product	Result Achievement	72,05		√		The achievement has been done from this program, either the counselor has done the identification of the students who has such a obstacle or give them follow-up
	Students' understanding	69,25		√		Based on the inventory results, it has good result.

Based on Table 4 it can be seen that the results of the evaluation of the product in the course of career guidance services in Vocational High Schools (SMK) in Banjarmasin to the achievement of execution program is 72.05% with good category and the students' understanding of their careers 69.25% in good categories.

5. Outcome Components/Summative

Based on the evaluation of the component Outcome/Summative can be presented in Table 5 below.

Table 5. The assessment of Components/Summative

Comp.	Aspect	Assessment Description			Conclusion
		Very Good	Good	Poor	
Outcome/Summative	The achievement of students' graduation	√			The achievement of students' graduation with 100% percentage in these 3 years.
	Alumni's ability to choose their job after graduation.			√	The school committee has no data about the alumni who has got job based on their passion.
	Alumni's ability to develop themselves.		√		The alumni have been able to develop their future based on their passion and their ability.

Based on table 5 it can be seen that the results of the evaluation outcome/summative program of career guidance services in Vocational High School (SMK) in Banjarmasin to the achievement of graduation target has gotten, and the result of the assessment is very good, and the ability of its alumni after graduation is they are able to choose and develop themselves / empower themselves according to their passion, and it is considered good.

IV. DISCUSSION

The results of the evaluation will be discussed based on the components Cippo/Summative.

A. Context Components

For the legal entity of the program's services has been fairly steady, according to UUD 1945. Vision-mission refers to the notion of education, both formulated in philosophical meaning, the Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional. However, the formulation of objectives not being implemented directly by all stakeholders. The collection of is done at the beginning of the new school year to new learners. But the results only reveal about matters of a general nature only, it is not specifically based on the characteristics of the expertise program.

B. Input Components

For the planning of career guidance, the program prepared is general, not specific on the nature, track, type and level of the department/study program at each school. Schedule and program activities in Vocational High School (SMK) in Banjarmasin has been planned in one school year, semester, monthly and weekly. The infrastructures tool which is existed in SMK Banjarmasin is adequate. Management of career guidance service program financing obtained from official sources and the responsibility of source of funding comes from the budget of the School Operational Assistance (BOS), the financing also comes from parents.

C. Process Components

The counselor has been carrying out activities in accordance with the schedule and plan career guidance service program, based on class assignment. Activities on career guidance service program service plans: Service orientation, information service, Cooperation has been performing well with the head and socialize with other stakeholders counseling. Monitoring guidance program of activities provided by the principal and teachers, but the monitoring results can not be proved by the results of a record / written report on the inputs provided.

D. Product Components

The achievement which done from this program is already well. Counselors have identified students who have barriers and have no inhibitions and are given immediate follow-up. For the career maturity of students obtained good results.

E. Outcomes/Summative Components

Data from the 2012/2013 academic year, 2013/2014, 2014/2015 recorded that SMK in Banjarmasin have reached the target of 100% graduation. School had already received the data alumnus working world according to their expertise Alumnus been able to develop a career direction accordance with the interests and talents (SMK 1, 4, 5). The alumnus have been able to develop a career direction in accordance with the interests and talents (SMKN 2 and SMKN 3).

V. CONCLUSION

The conclusion that can be derived from the context is : the evaluation input either category/ adequate, the evaluation process is viewed from the five aspects of the implementation process which is to execute the activities in accordance with RPL either category, delivery of materials both categories, media services: poor category, the service method: poor category and for process monitoring and cooperation with school stakeholders is good The evaluation of category of products for the achievement and students' understanding of career categories. Evaluation of outcomes / summative for the achievement of the target graduation is very good ratings and the ability of its alumni after graduation is they are not really able to choose a career and the ability to develop self alumni / empower themselves according to their passion.

REFERENCES

- [1] Furqon dan Aip Badrujaman.(2014). Model Evaluasi Layanan Dasar Berorientasi Akuntabilitas. Jakarta: Indeks.
- [2] Matthew B. Miles dan A. Michael Huberman.(1994). Qualitative Data Analysis: an Expanded Sourcebook.California: Sage Publication.
- [3] Munandir.(1996). Program Bimbingan Karir di Sekolah. Jakarta: Depdikbud.
- [4] Ruslan A. Gani. (2001). Bimbingan Karir. Bandung: Angkasa.
- [5] Robert O. Brinkerhoff. (1986). Program Evaluation: A Practitioner's Guide For Trainers and Educators. Boston: Kluwer-Nijhoff Publishing.
- [6] Sugiyono. (2009). Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R & D. Bandung: Alfabeta.
- [7] W.S. Winkel dan M.M Sri Hastuti. (2006). Bimbingan dan Konseling di Institusi Pendidikan. Yogyakarta: Media Abadi.

CIPP EVALUATION OF THE LEARNING IN CULTURAL DIALOGUE DURING UNSOED INTERCULTURAL SUMMER-CAMP

Oscar Ndayizeye, Agrégé TEFL

Burundi Higher Teacher Training School (Ecole Normale Supérieure, ENS)
oscarndayiz@gmail.com, ndaosca@yahoo.fr

Abstract - This research intended to: (1) evaluate how the Unsoed Intercultural Summer-camp's (UIS) 2015 culture learning context, input, process and product were; and (2) draw conclusion whether the UIS 2015 was a real intercultural situation. In this descriptive and qualitative CIPP evaluation, the saturated sampling technique was used due to small sample size, the camp participants and their guides/tutors (n=27) were both the population and sample. The data were collected through researcher's complete participant observation, checklist, document analysis, conversational interviews and Group Evaluation Forms (GEFs) instruments to ensure truth worthiness and data, sources, and perspective triangulations. Data analysis technique was "the data reduction-Display-Conclusion" one. After comparing UIS 2015 goal fulfilment indicators to pre-set criteria, the findings in this study show that all the indicators made from the UIS 2015 program highlights were fulfilled, GEFs showed that the participants partook in all the activities on the menu, and conversational interviews indicated the participants' satisfaction of the camp organization. The occurrence of faux-pas, English language use problems and participants intercultural presentations all crown the UIS 2015 as having been a real intercultural meeting point and a successful program.

Keywords: *CIPP evaluation model, program evaluation, cultural dialogue, intercultural communication*

I. INTRODUCTION

The current era is a core around which spin globalization, information technology and communication, and internationalization. All these are there to collapse barriers that used to prevent people from communicating internationally. To destroy such former barriers, availability for partnership and frank collaboration have to be each nation's equipment.

In this perspective, the younger generation, who are the active actors in the globalization era, should know about international experiences. As an instance, the natural disasters rescuing operations, since they require an international commitment, require people with multicultural and multinational experiences to intervene, Baldwin, R. et al. (2014:3) [1].

Such experiences are needed as skeleton keys for the young. For example, while trying to increase the South East Europe youth action, SALTO South East Europe Resource Centre (2011: 19) [2] ascribed, on the theme of "Dialogue Promotion," the objective of "*dialogue and collaboration amongst individuals belonging to multi-cultural and multi-religious communities. A further aim was to increase participants' awareness and skills in working with issues of religious diversity in multi-religious and multi-cultural environments.*" This is a patent example tokening how international experiences are needed and thus shared through intercultural or youth camps.

A. *Knowing about other cultures: a necessity*

Culture has been defined by so many but a scholar to the extent that it is hard to give a definition which is likely to satisfy everyone. But the definition that is worth mentioning here is that

culture simply refers to the way of life of a group of people, the symbols they use, values, behaviors, artefacts, and other shared aspects, Baldwin, R. et al. (ibid: 5) [1]. Griswold (2013: 135) [3] seems to give the rationale to studying about other cultures. He comments that if culture really involves shared meanings, to move in different cultures might require understanding different systems of meaning and assumptions, principles, and nuances that any particular cultural object may evoke in such systems. The young generation, which is coping with vagueness about globalization, needs then to know about other cultures.

B. Program Evaluation

Under this heading, some terms need clarifying. To begin, Stufflebeam and Coryn (2014: 312) [4] give an operational definition of evaluation as the process of delineating, obtaining, reporting, and applying descriptive and judgmental information about an object's value, as defined by such criteria as quality, worth, probity, equity, feasibility, cost, efficiency, safety, and significance. Note that program is a set of resources and activities directed toward one or more common goals, especially under the direction of a single manager or management team, Wholey, Kathryn, and Newcomer (ibid: 652) [5].

C. CIPP Evaluation Model

CIPP refers to Context, Input, Process, and Product. The CIPP Model is in line with a systems view of education and human service. Karimnia and Kay (2015:87-88) [6] detail about CIPP:

Context evaluation: What needs to be done? It assesses needs, assets, and problems within a defined environment; Input evaluation: How should it be done? It assesses competing strategies and the work plans and budgets of the selected approach; Process evaluations ask: Is it being done? They monitor, document, and assess program activities; Product evaluation: Did it succeed?

D. Research problem

Intercultural Summer Camps are planned and carried, but the literature shows literal about how they can be evaluated formally. For the case at hand, a series of questions may be put forward:

1. How were the context, input, and process of UIS 2015?
2. How did the UIS 2015 reflect a real intercultural situation?
3. To which extent was the UIS 2015 successful?

E. Unsoed Intercultural Summer-Camp 2015

Unsoed Summer Camp (UIS) is a student exchange program offered by Jenderal Soedirman University (Unsoed). UIS provides international students the opportunity to learn not only intercultural experiences but also integrated understanding on local wisdom, ecology and agriculture of Banyumas Regency, a beautiful Indonesian countryside in Central Java through a valuable and meaningful seven day program, IRO/Unsoed [11]

II. METHODOLOGY

A. Determining the UIS 2015 success

Based on the UIS 2015 program highlights, the holistic-criteria based rubric used to determine the success of the UIS scores the very program indicators, i.e whether participants were:

- 1 a. Taken to batik making workshop
- 1 b. Watched / Taken to Javanese dance session
- 1 c. Taken to gamelan ensemble,
- 1 d. Taken a local culinary (nopia) making house
2. Taken to a mangrove and coastal Excursion
- 3 a. Taken to Palm Sugar making place
- 3 b. Taken to a paddy field
4. Tropical Forest Adventure
5. Make Intercultural Presentations
6. Introduced to Independence Day Events

Note that indicators 3 b. and 6 are excluded because the UIS 2015 organizer judged better not to carry them due to constraints the researcher did not manage to find out. The success in

terms of percentage is computed like: (the received score: Max. Score) x 100 on the score stated as follows:

Table 1. Determining UIS 2015 success

indicator fulfilment	The score
Each indicator	1
The Max. score	8

In this study, the quantitative criteria for the effectiveness of the UIS 2015 is determined, for indicators whose fulfilment is verified through checklist and Group Evaluation Forms (GEFs). The success criteria are determined by ninety percent (90%) of Yes on the checklist, Wilson (2013: 25) [7].

B. Research design, Population, and Sample

The research design is a descriptive qualitative program evaluation research. In this study, the population is made of the eighteen (18) UIS 2015 participants and nine (9) regular guides, all of them making n=27 persons. The sampling method used is a non-probability one, and since the population is less than thirty participants, the sampling technique used is a saturated one; the sample is equal to population, Sugiyono (2015: 124) [8].

C. Data collection

1. The source of data:

The sources of data in this study are the participants (both countries delegate students and their local guides) and documents related to the preparations and ongoing of UIS 2015.

2. Instrumentation:

Three data collection instruments are used in this study: (1) Participant observation's GEFs and checklist sheets; (2) conversational interview; and (3) document analysis/review. Note that observation is "*the systematic description of events, behaviors, and art efacts in the socialsetting chosen for study,*" Sugiyono (ibid: 310-311) [8]. The GEFs and checklist observation sheets, together with document review, yielded data about both the context, input, process, process, and product evaluations. But the product evaluation is also carried through conversational interview (where the interviewer reacts as well as shares to create a sense of conversation) and comparing the indicators to the criteria.

3. Data analysis technique

The analysis of data is a qualitative one, especially the Miles and Huberman (1984) qualitative data analysis model, Sugiyono (ibid: 337) [8]. Such a model consists of (1) data reduction, i.e to make them appear simple for easy analysis; (2) data display which summarizes data after reduction, sorts them by discarding irrelevant data: this will turn the data into ones which are easily understandable, and then (3) Conclusion drawing/verification.

The analysis of the checklist data and that from GEFs, Watson (2014: 215-216) [10], is done through *Comparison Against Criteria*, Wilson (ibid: 25) [7]:

In this technique, you compare the results of a checklist against a quantitative criterion. So, you might, for example, compare the percentage of Yes answers to a criterion value that 90% of the answers must have a Yes answer

D. Data Analysis, Findings and Discussion

To evaluate the UIS 2015 program via CIPP model, the effectiveness the program depends on how both the context, input, and process of the program are. The checklist consists of ticking Yes or No following whether the elements are present or no. The scrutiny of available documents makes the basis of the checklist data for:

Table 2. CIPP Evaluations

1. Context Evaluation (concerned with Camp objectives and activities)

1. The objectives ascribe learning about intercultural experiences
2. The objectives announce a wide range of knowledge that will be acquired
3. The objectives are menus of how campers will be introduced to Indonesian culture
4. The objectives anticipate a sort problem solver knowledge during the intercultural camp
5. The camp activities reflect the camp program highlights
6. The camp agenda is socialized to the campers
7. There is no activity improvisation
8. The camp activities are loaded with intercultural situations

2. Input

9. There are booklets for participants
10. Unsoed Presenters' profile: Presenters are specialists in the subjects they present
11. Experts presentation materials are displayed, detailing aims of the presentations and printed visual media are clear
12. Projector
13. Computer-Assisted facilities
14. (Soft) Board and markers
15. Artefacts, cultural tools/objects are availed for the sake of concreteness
16. No complaints about Hotel/homestay rooms not meeting the minimum ease and sanitary requirements

3. Process (During Cultural Guidance & Experts' presentations)

17. The experts explain their materials clearly and give feedback to participants' questions
18. Topics presented by the experts are ascribed on the program highlights
19. Experts respect time allocated to their presentations
20. No activities overlap with others and there are none of them cancelled due to time constraints
21. Participants are given opportunities to handcraft some cultural object guided the skilled guide/tutor

1. Product

22. The campers exchange their countries' cultural tokens/greeting ways orally and/or in the written form
23. The campers are given opportunity to share/perform formally some of their countries' cultural aspects.

a. Data from Group Evaluation Forms (GEFs)

Based on the participants' presence check every day before any activity started, the following is the Group Evaluation Form showing whether a participant from X country has participated in

the offered activities A_s ; and this is another way to assess the UIS 2015 participants' fulfilment of the summer camp goal indicators:

Table 3. GEFs Data Description

Country Participants		UIS 2015 Goal Indicators								
		1a.	1 b.	1 c.	1 d.	2.	3	4	5	
Burundi	B1	X	X	X	X	X	X	X	X	
	B2	X	X	X	X	X	X	X	X	
China	C1	X	X	X	X	X	X	X	X	
	C2	X	X	X	X	X	X	X	X	
	C3	X	X	X	X	X	X	X	X	
	C4	X	X	X	X	X	X	X	X	
Japan	J1	X	X	X	X	X	X	X	X	
	J2	X	X	X	X	X	X	X	X	
Malaysia	M1	X	X	X	X	X	X	X	X	
	M2	X	X	X	X	X	X	X	X	
	M3	X	X	X	X	X	X	X	X	
Papua N. Guinea	PN1	X	X	X	X	X	X	X	X	
	PN2	X	X	X	X	X	X	X	X	
Solomon Islands	S1	X	X	X	X	X	X	X	X	

b. Data discussion

As criteria set for checklist for UIS 2015 to be effective amount to "90% of Yes," the context, input, process and product evaluations mirror the following observations about the checklist data:

Table 4. CIPP Evaluation Results

Evaluation	Yes	Score	%
1. Context	8	8	100
2. Input	8	8	100
3. Process	5	5	100
4. Product	2	2	100

The sum of scores gives 23 per 23 and once converted to percentage, it yields 100%. The set criteria for each evaluation of the four (CIPP) being 90% achieved, this means the criteria-score has been achieved. So according to the data from the checklist, the UIS 2015 has been effective.

Concerning data from the GEFs, if we limit to delegates from abroad (since Indonesian guide knew every cultural aspects that UIS 2015 participants were being exposed to), they participated in all the activities presented (see table 3 above). Eight activities of eight leading the success indicators are fulfilled:

Table 5. GEFs Results

Activities	Score	Success in %
8	8 out 8	100

So since all the participants partook in all activities exposed to them, the eight indicators were attained hence the criteria "of 90% participation in all activities" being exceedingly attained.

The analysis of data from the "Conversational Interview and recorded Quotations or Facts" indicates support to the UIS 2015 success rate but also presents the weaknesses of the program and analysis

- 1) The UIS 2015 is a really situation of intercultural communication, that required participants to make use of intercultural think to maintain the ongoing camp activities. The explanation about the rice and Wayang legacies, why the Banyumas headquarters is now in Purwokerto, etc revealed a local wisdom and here are more elements tokening how the situation was truly that of intercultural dialogue where we had:

-
- a) Faux pas (Cummings, 2009:129): some miscommunication were attenuated like this: *"This is the last chance for you to eat... (two bowls)!"* "My last chance?"
 - b) >>>There was a great confusion for the hearer (from Japapan). But he contained his confusion and asked to the local guide (speaker) "My last chance?" and a further explanation was given bringing back common understanding.
 - c) Cultural shock: A Solomon Islander narrated her missing home food during her first day in Indonesia
 - d) An Unsoed official intended to handshake Malay she-participants, and they did not shake his hand due to religious grounds, which discomforted him a bit.
 - e) English language was a restrictive matter to most participants. Although it was among eligibility criteria "Students who are able to communicate in English," shortcomings due to lack of that communicating medium became a spot mark: this caused subgroups amongst the participants; which even created preferences:
 - The Thai preferred always to be with their fellow Thai or at least with Indonesians who could understand Malay (for South Thailand side speaks Malay), a language close to the Indonesian one.
 - The same case was observed among Chinese where only one could speak English and then was an interpreter for the rest of the Chinese delegates.
 - The interpreter at the experimental farm deter the communication, turning it into a noisy room due to her failures to interpret cow farming terms as an expert that speaks only Indonesian was presenting his material.
- 2) We have also views sanctioning the success of the UIS 2015:
- The Burundian participant attesting that there are palm trees in his country but not used to make sugar shows a great lesson he has learned from the camp
 - Answers like UIS 2015 is "very nice," "cool, very interesting" are token of satisfaction, good indicators of "the product evaluation" in this study.
 - Another product evaluation success indication is that ceramic crafts that participants made themselves and palm sugar were eagerly received by the participants, and the researcher managed to hear a participant releasing: *"Good...I'll go home with it...it will be my souvenir about Indonesia"*
 - The last tokens sanctioning the success of UIS 2015 as a real intercultural dialogue are the intercultural presentations where each participant delegation presented some cultural elements of their own countries.

III. CONCLUSION AND SUGGESTIONS

According to the CIPP evaluation of UIS 2015 program, the program was successful: the objectives and activities (context) are clear and were clearly implemented; the camp learning resources and facilities (input) are availed and they were effective and informative. The participants also were given opportunities to handcraft, to handle some Indonesian cultural objects (ceramic, Wayang, nopia and batik making) which is an indicator of a program with a fruitful process. And since every participant participated in all the activities fulfilling at 100% the UIS 2015 indicators, the exposure to local culture and sharing of the delegates' countries' cultural aspects, occurrence of faux-pas among participants, stereotypes, all together with the satisfaction expressed in conversational interviews both crown the UIS 2015 as having been a real intercultural situation and a successful program.

For other researchers in Indonesian context, there should be developed a valid instrument to assess cultural camp participants' satisfaction. It was not used here because the researcher felt that using a tentative and no validated one would not be fair and ethical as it could over esteem or jeopardize at least the image of Banyumas' cultural legacy.

Few UIS 2015 pitfalls were located on the realm of English language: it was a requirement to attend the summer camp, but most participants were object to stereotypes from their co-campers as not willing to associate with others.

As suggestions, the interpreters in forthcoming UIS should be initiated with technical English they intend to facilitate in: experimental farming, etc based on the activities UIS organizers intend to expose to their camp participants. Organizers should consider little use of English language by participants makes subgroups among campers and this dilute the UIS an international meeting point.

REFERENCES

- [1] Baldwin, R. J, Coleman R. M, Gonzalez A, & Shenoy-Packer, S. *Intercultural Communication for Everyday Life*. West Sussex: Willey Blackwell, 2014
- [2] SALTO South East Europe Resource Centre. (2011). www.salto-youth.net/downloads (retrieved on Wednesday, 20 January 2016 at 12:39)
- [3] Griswold, W. *Cultures and Societies in a Changing World* (4ed). Los Angeles: Sage, 2013.
- [4] Stufflebeam, D. L., & Coryn, C. L. S. *Evaluation, Theory, Model*. 2014.
- [5] Wholey, J. S., Hatry H. P., Newcomer, E. K. *Handbook of Practical Program Evaluation*. CA: Jossey-Bass, 2010
- [6] Study, C., Karimnia, A., & Kay, E. "An Evaluation of the Undergraduate TEFL Program in Iran," *International Journal of Instruction*. Vol.8, No.2, July 2015, pp. 84-98
- [7] Wilson, C. *Credible Checklists and Quality Questionnaires Credible Checklists and Quality Questionnaires A User-Centered Design Method*. Amsterdam: Morgan Kaufmann, 2013
- [8] Sugiyono. *Metode Penelitian Pendidikan. Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung : Alfabeta, 2015.
- [9] Cummings, L. *Clinical Pragmatics*. Cambridge: Cambridge University Press, 2009
- [10] Watson, J. *Intercultural learning and development among youth participants in the short term educational programmes of an international charity (CISV)*. Unpublished PhD thesis, Birkbeck, University of London, (2014)
- [11] <http://iro.unsoed.ac.id/content/uis-unsoed-intercultural-summer-camp-2015>

EVALUATING BASIC ENGLISH TEST ITEMS FOR NON-ENGLISH STUDENTS FROM TEACHERS PERSPECTIVES

Prihantoro

Universitas Diponegoro
prihantoro2001@yahoo.com

Abstract - This paper seeks to evaluate diverse Basic English tests items for Indonesian college students who are non-English major (all majors except English). The object of this research is a collection of English exam sheets (both mid and final terms) for non-English students in one state university in Central Java, Indonesia. Test items were categorized thematically. The most popular skills tested are reading comprehension and grammar. Writing is low. Listening is never tested. In order to understand the perspectives the English teachers, the teachers who composed the exam sheets were also interviewed. The result has shown that 95% of the items are indirect assessment (multiple choices, true false, matching, cloze quiz) are preferred as the correcting and scoring is much easier. Besides, this gives advantage for students as the answers are already provided. It fits language skills like reading and grammar. Direct assessment in the form of construction (e.g essay) is rare. The existing ones put very tight constraints (number of words, topic etc). Direct assessment is less preferred as correct/proper answers may vary and the scoring is time consuming. The interviews result show that classes they teach and their teaching loads drove this preference. However, changes might take place when the related program/department/university/ 1) strengthen the human resources (in terms of quantity and quality) or 2) delegates the responsibility to a committed language center.

Keywords: *direct and indirect assessment, teacher perspectives, large classes*

I. INTRODUCTION

This paper looks into test items in Basic English class, a mandatory class for all college students (except English Department students) in Indonesia. Creating test items for this course is considered challenging. First, in terms of entry criteria, this course can be taken without any requirement; where students have different degree of English language proficiency. Second, it is a mandatory class that students must take, even though it may not have a direct link to the core competence of the department; they have to take it whether they like it or not. Three, classes are usually large in size (more than 40 students per class).

With these facts, it is important to look at different aspects of this course, and suggests how to improve it. This paper discusses the test items designed by the teachers. Besides the evaluating the items, it is also important to understand how teachers reflect on the items; what motivate such items to surface; what has been considered advantage or disadvantages of the items.

This paper is organized as follow. Section 1 describes the background of this study. Section 2 briefly visits some basic concepts in language testing which are discussed in this study. Section 3 describes methodology. Findings are discussed in section 4, while section 5 proposes some possible solutions.

II. LANGUAGE TESTING

There are many definitions of language testing, especially when we test specific skills like reading, conversation, listening, grammar, translation, writing and etc. However, we are aware what testing is. It is an instrument to measure something. As quoted from Douglas (2009) language testing is an instrument to measure language ability.

There are at least two kinds of testing, formal and informal (Popham & Popham, 2005). Informal testing is also an important part of testing, but done informally; like when teachers ask questions to students to measure their understanding of what has been taught or just to discover their background knowledge. In fact, informal assessment is a part of learning process.

In this paper, however, we focus more on formal assessment (Fulcher, Practical Language Assessment, 2010). Formal assessment is important as it gives standards, in form of grades, which (should) become the reflection of the learning process. With the results of formal assessment, one who is not involved in the learning and teaching process may understand one's competence.

A. Why, What and How

Before discussing the form of a formal testing, it is important to understand the reasons why the testing is given (Fulcher & Davidson, 2007). Consider diagnostic test (Bachman, 1990). This test is used to measure one's competence and then decide whether the competence is up to standard or not. As an illustration, students registering for an International class must take written and oral test in English, as English will be the language of instruction. It is important for the students to pass the test as they are going to write reports, discuss and argue in the class by using. Another illustration is final term exam, for a business English class as an instance, where the reason for testing is to check whether learning objectives are achieved. These two illustrations are example to show the importance of reasons for conducting a test.

When the reasons for conducting a test have been established well, the next step is to identify what to be tested. (Hent, 1987) Is it the acquisition of certain grammatical structures (passive, causative, clauses, parts of speech)? Is it to measure vocabulary acquisition in certain domain (business, academic, sports)? Is it measure fluency in conversation (use of fillers, controlled ideas, flow, and responsiveness)? What competence to measure is usually present in learning outcomes (as a part of lesson plans).

After identification of what to be tested is cleared, it is important to set how the test should be organized (Fulcher & Davidson, Language Testing and Assessment, 2007). Does the test have to be held in oral? Should the oral test be divided into several segments (introduction, short talk, longer interactive talk)? Does it have to be written? What form of test items are appropriate (true-false, multiple choice, essay)? Answering the question of why, and what, is very important to determine the 'how'; which is how we execute the test.

B. Direct and Indirect Assessment

Another type of assessment is direct and indirect assessment (Breland & Judith, 1979). This will be the core of the discussion in this paper. However, let us briefly review shortly the difference between the two types of assessments. Direct assessment is a type of open testing, like essay or interview. Responses required in this type of testing are open, which means it is not prepared by the teachers or test designers. Therefore, answers may vary, and there can be no exact answers. What teachers can do is measure whether the answers (to what extent) fit the variables of the tests from rubrics that have been prepared.

Indirect assessment is a type of testing where selected responses may be provided by the teachers or test designers. This includes true-false identification, matching sequence, or multiple choices. In this paper, we focus on indirect assessment, how teachers reflect on this in Basic English class, and the motivation to prefer this assessment as compared to indirect assessment.

III. METHODOLOGY

Data in this paper is obtained by using questionnaires where the respondents are from one state university in Central Java. The questionnaire will be discussed in the discussion section. I requested the list of English teachers from different programs/faculties and randomly sent requests for their participation in this research. I do not limit their status (permanent or part time). I also do not distinguish educational background (linguistics/literature/area studies/ Non-English departments). Those who have agreed to participate, received questionnaires. From 15 questionnaires, 13 are returned. One questionnaire is not returned because the respondent had to perform religious travelling, and another one was not returned for any reason. As the questionnaires were filled by handwriting, respondents are visited on the purpose of validation. Besides validating handwriting, another purpose of the visit is to request further clarification (when needed). The data in the questionnaires will be presented by both visual chart and prose description when necessary.

IV. DISCUSSION

One of the main findings in this study is that the number of indirect assessment is weigh over direct assessment. The percentage of indirect assessment test items is 95% in total with 60% in multiple choices. Subjects tested by using this test items are mostly reading and grammar.

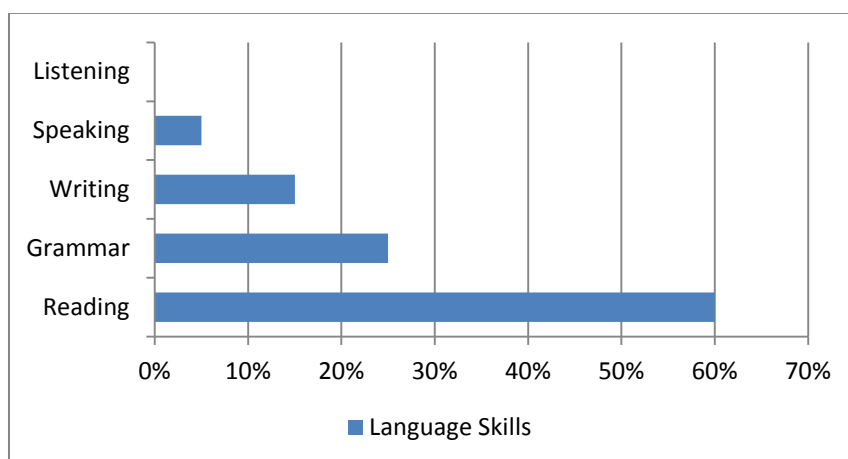


Figure 1. Distribution of Language Skills Tested

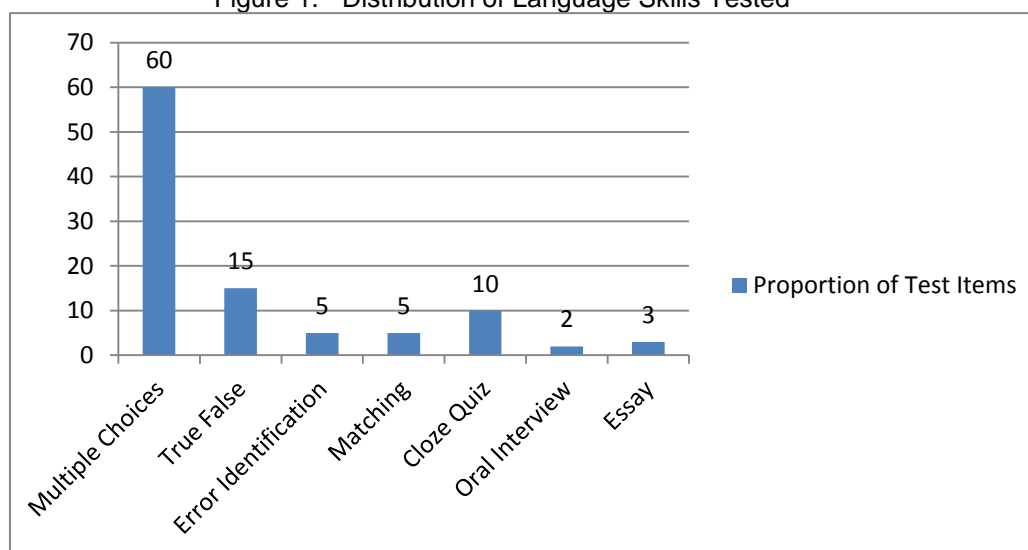


Figure 2. Proportion (%) of Test Items

Reading and grammars dominate language skills tested in the exam. According to the respondents, these two are selected as they are usually the objectives of the course, and they are skills realistically to be tested considering the size of the class. And what has been said about the preference of multiple choice?

Teachers have their practical considerations why this kind of selected response item is preferred (Roediger & Marsh, 2005). First, it reduces the anxiety of the students. Teachers believe that during direct assessment such as essay writing or oral interview, students feel anxious and as a result cannot perform well. Second reason is about the scoring. Scoring has always been in the mind of the teachers or test developers when they design tests. I survey how many credit hours teachers have to teach, and none of them teach below 10 credit hours (1 credit hour = 50-60 minutes). Most of the teachers teach 10-20 credit hours, but some of them teach more than 20 credit hours per week.

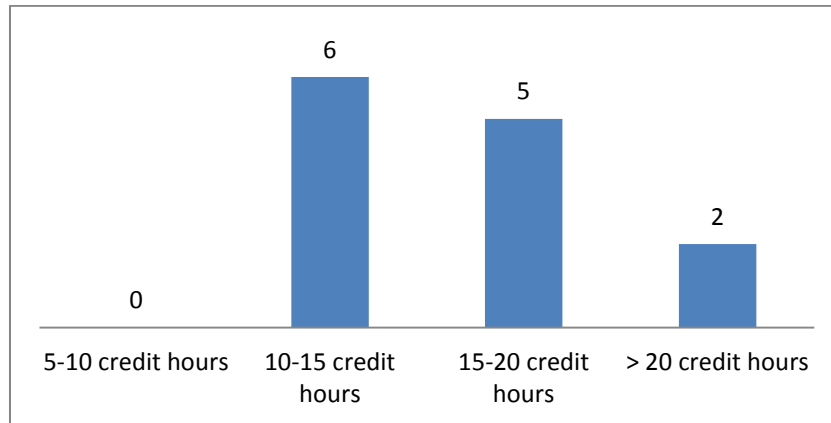


Figure 3. Teaching Loads per Week

I also ask them about their status. Three of them are part timers and the rests are permanent civil servants teaching in the University. Both are equally problematic. While the part timers also have other (teaching) job in different institutions, those who are in permanent positions are obliged to publish and also to do community service. Publication usually is the most challenging, as compared to teaching and community service. Often one has to spend months to publish in a national journal, as papers may have to go through revisions. With these teaching loads, indirect assessment is a logical choice. Scoring can be delegated to other people, as there is absolute answer within the choices.

Consider also the number of students in a Basic English class, which according to the survey mostly is within 50 to 100. Even the respondents mention that there are some classes where the students are beyond 100. Multiply this number with scoring time. If an answer sheet consisting the combination of direct and indirect assessment is scored within 15 minutes, finishing 100 sheets will take 1500 minutes or 25 hours.

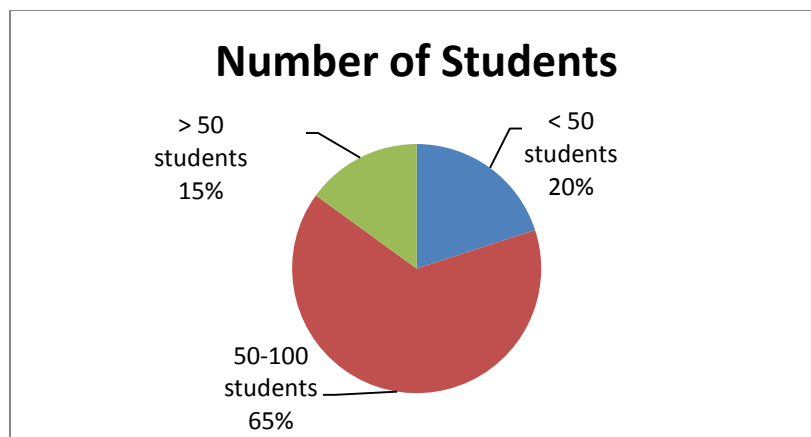


Figure 4. Class Size

A large class is an issue (Bachman, 1990) here; why in a foreign language class, the number of students may reach more than 100. Some teachers talk to the authority (the department or faculty member), and reasons are complex, but to some extent, make sense. Some mention that relates to the financing. If they have to divide 100 students into three or four class, then the authority would have to think to hire more teachers. This also happens to other non-department class. As an illustration, take the example of civil engineering department where students in the earliest semester must take basic science classes like basic math, or physics. So, the issue is financing, which really makes sense for faculties or department with tight budgets. Another reason is room availability. In some departments, they only have rooms for large classes. It takes time to build rooms for language classes with 40 students, as they have to propose this first to the university. Even in some departments, scheduling is very tight as number of classes are very

close to number of subjects. When a class is divided into two or three, they might be short of classroom.

When asked about the type of text items that they would use (assuming that classes are in the ideal condition from teaching loads number of students to facilities), teachers prefer the 70%-30% combination of indirect and direct assessments. Improvement in direct assessment is required to measure production skills like writing or speaking, while indirect assessment will fit best for understanding like reading or grammar.

Indirect assessment test items are also challenging in some aspects. First, it takes more time to prepare, as teachers or test developers would have to come with different alternatives as in the case of multiple choice. Second, the alternatives (or called distractors), are also not easy to find. They cannot be too different, but cannot also be too similar; achieving reliability in this kind item is not easy. Sometimes, test items need to be tested too before they are distributed to students. Instruction must be set very clearly in order not to confuse students. So, besides the scoring efficiency time benefit for teachers and selected responses benefit for the students, multiple choices item is challenging in terms of reliability.

According to the teachers, these all are classical problems, and they, and also the faculty or department members are actually aware of this. But why these classic problems do not end? One answer is because in many department, (or all departments) Basic English is not the core competence. This is a very logical concerns. Basic English falls to the type of *Mata Kuliah Umum* (MKU) which is general subject in literal translation. Some other subjects that fall to MKU are subjects like sports, religion, or Pancasila (five pillars of Indonesian statehood). They do not directly contribute to core competence of the departments (physics, chemistry, engineering, psychology), but students cannot graduate without passing these courses. Why not removing these courses from the curriculum when they do not have major contribution to core competence? Why not replacing them with more practical course such as entrepreneurship and internship?

It is not that easy regarding the impacts it may probably cause. Permanent teachers might have problems with the removal if MKU courses are the only courses they rely on. In the university where the questionnaires are distributed, permanent teachers are recruited from English department, so when such courses are removed they can still teach in English department. But consider this, not all university have English department. When such courses are removed, the permanent teachers will have difficulty in meeting the minimum threshold of teaching loads. Where should they teach? Then why not making this subject elective? This is actually a good idea, but quite challenging for implementation. Who can guarantee that students want to take this? Basic English is usually only offered without further progress (intermediate, advanced English), so that they will have to go outside to acquire English. There are so many English language institute out of the universities provide more interesting offers for English language development. If this subject is elective, why not starting from the scratch outside of the university? Unless the offer (teaching methods, facility, teaching aids, curriculum) is better, students will definitely choose to learn outside.

V. CONCLUSION

I here offer two solutions, but I totally understand that in reality, they are really difficult to implement and can only be successful with the support of students, teachers and including budgets and policies from higher authorities. In a university with English department, it might not be a good idea to request them to teach Basic English. It is better to recruit their own teachers with specification on TESOL and sub-specialization in English for Specific Purposes. This has actually been done in some universities; let's say an English teacher who works under the department of agriculture. But the problem is, the teacher is also requested to teach in other department. It is better that an English teacher in college focus on only one department with 12 credit hours max. In this way, the direction of research, publication, teaching and test material developments will focus on the nature of the department. If one stay consistent under one department, say department of geology, s/he will have time to develop teaching and test materials specific to geology. There will be enough time to research geological texts to feed those materials. One can also develop a teaching method and write papers, and publish on teaching English under geological domain. The department should assist the teacher by providing assistance such as teaching aids, access to geological lab, involving in the curriculum design and etc.

Another solution is to request the help of the language center. Each university usually has a language center. Ideally, teachers working in the language center should be recruited by the university, not department. However, university may delegate the teachers to different department, but with careful considerations. Teachers are recruited on the basis of TESOL and English for Specific Purpose, as it has been commented previously. Therefore, when recruiting teachers from literature or area studies (common sub-specifications in English Departments), they must be provided with further training. In this way, the oversee of the teachers is under the authority of the language center. Therefore, the language center has to be excellent. It is not enough for the language center to have resources only for conversations or international tests like TOEFL or IELTS. They must have resources from many different disciplines to assist teachers design teaching and test materials. In my personal opinion, permanent position should only attach to language center managements, while for the teachers it is better to hire English teachers on yearly contract basis. There is a dilemma for state universities. It is quite difficult these days for state universities to propose a new position to the government; therefore teachers in language centers are often recruited from inside circle in the university with remuneration besides they basic salary. It is not good for their career development. Too many teachings will affect publication, and not all publications are related to teaching. Some have expertise on literatures, some others on linguistics, some others on area studies, other might be on translation and etc. But recruiting teachers might be related to budgeting, where some universities have different ways in caring for their language center.

REFERENCE

- [1] Bachman, L.-F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- [2] Breland, H.-M., & Judith, L.-G. (1979). A comparison of direct and indirect assessments of writing skill. *Journal of Educational Measurement* 16.2, 119-128.
- [3] Douglas, D. (2009). *Understanding Language Testing*. Routledge: London and New York.
- [4] Fulcher, G. (2010). *Practical Language Assessment*. London: Hodder Education.
- [5] Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. London and New York: Routledge.
- [6] Hent, G. (1987). *A Guide to Language Learning: Development, Evaluation and Research*. Wadsworth: Foreign Language Teaching Press.
- [7] Popham, W.-J., & Popham, J.-W. (2005). *Classroom assessment: What teachers need to know*. New York: Pearson/Allyn and Bacon.
- [8] Roediger, H. L., & Marsh, E.-J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(5), 1155.

Is the German language text too short for the senior high school students?

Ryan Nuansa Dirga¹, Primardiana Hermilia Wijayati²

¹Study Program of Language Teaching, Graduate Program, State University of Malang

²Literature Faculty, State University of Malang

ryandirga@gmail.com, primardiana.hermilia.fs@um.ac.id

Abstract - This article discusses the instrument to measure reading skill of high school students in Malang. The instrument is developed based on the curriculum 2013. There are some kinds of texts in the test that have different word amount. The short texts consist of 20 until 47 words and the students need 30 until 45 seconds to understand each text. Besides, the moderate texts consist of 57 until 82 texts. The students need a minutes to understand each text. The long texts consist of 114 until 126 words. The students need two minutes to understand each text. The result shows that the students need different time to understand the texts.

Keywords: curriculum 2013, German language, instrument, reading

I. INTRODUCTION

Curriculum plays an important role in the education system. A curriculum is a plan and regulation system that contains learning goals (competences), content, learning materials, and implemented learning methods. Curriculum is used as the foundation of the education system in order to achieve the national education goals [1]. Curriculum in the education system of Indonesia has changed from time to time. The curriculums that have been applied in Indonesia were curriculum 1994, 2004, 2006 and the newest curriculum is the curriculum 2013.

The curriculum 2013 is not a really new curriculum. It is a development of the curriculum 2006. The curriculum 2006 has some weaknesses, so it is improved. The improved aspects are the content, the competences, learning process, and the assessment [2]. The content of curriculum 2006 is too full. We can see through the amount of the subjects. There were too many subjects. In curriculum 2013, some subjects are integrated. Most of the competences in curriculum 2006 are emphasized on cognitive aspects (knowledge). In curriculum 2013 the basic competences are emphasized not only on cognitive aspect but also on affective (attitude) and psychomotor (skills). In the education aspect, there is character education. The achievement of the students is assessed from their competence. Besides those differences, there is still a difference in the implementation of curriculum 2013. In the implementation of curriculum 2006, teachers need to produce a syllabus. Then they develop the syllabus in the learning plan. In the curriculum 2013 that is different.

In the implementation of curriculum 2013, teachers do not need to produce syllabus, because it is already available with the curriculum. In the syllabus there are four core competences (*KI*) [3]. Those are spiritual competence (*KI* 1), social competence (*KI* 2), cognitive competence (knowledge) (*KI* 3) and psychomotor competences (skill) (*KI* 4). These four core competences are elaborated in basic competences. The basic competences are taught in a learning process.

At the end of the learning process, the assessment should be done. Assessment is a process to know whether the learning goals are achieved or not [4]. Based on the interview with German language teachers, they stated that they have difficulty in writing test based on the curriculum 2013. The reason is because the curriculum 2013 is new. This situation affects the quality of test.

Teachers made test for German language subject does not have good quality [5]. Although it has good reliability, but the content validity of the test is low. The items of the test are mainly easy and moderate. There is no difficult item. The most items are in cognitive level A1 (remembering). The distractors are not functioning. In conclusion the test has low quality.

Because of the test's low quality, the test can't measure the students' performance correctly. Furthermore, it is difficult to know whether the students achieve the required competences or not and whether the learning goals are achieved or not. Developing a good test is really important in order to know the achievement of the students. Teachers should know how to develop a good test. A good test has some characteristics.

We can see through some aspects to know the quality of a test. The characteristics of a good test are validity, reliability, objectivity, practicability, and economic value [6]. The other characteristics of a good test are the quality of the items [7]. A good test should be developed for each language skill. This paper discusses the development of a test instrument to measure students' reading skill. The test instrument is developed based on curriculum 2013.

II. RESEARCH METHOD

This research is a development research. In developing the test instrument, the development model of Cennamo and Kalk is used [8]. Research and development is a process or procedures to develop a new product in order to improved the older product or to prove the efficiency of the product [9][10]. A product can be produced with these procedures.

The developed products are not only hardware e.g. books, modules, learning mediums, but also software e.g. computer programs or learning model. In this research, the developed product is a test for German language subject. There are some procedures that should be done to develop the test. These procedures are analyzing the needs, determining the blueprint, developing the prototype and trial [8]. The trial aims to know the quality of the test.

In this research, there are five stages that have been done. These stages are from the development model of Cennamo and Kalk. The stages are define, design, demonstrate, develop, and deliver.

A. Define

In the define stage, the researcher identified the characteristics of the students. The need is a test for German language subject based on the curriculum 2013. The students are the students of the 11th grade class language program and *Lintas Mlnat* program. Then the researcher determined the product, namely test for German language subject, and the materials for the test. The materials are materials for German language subject for the 11th grade class language program with *Alltagsleben* (daily life) theme. In this stage, the researcher has determined the item's form, namely multiple-choice. There are five distractors. The test aims to measure the reading skill of the students.

B. Design

In this stage, the researcher determined the sub themes for the test. The theme *Alltagsleben* has three sub themes, namely *Essen und Trinken* (food and drink), *Kleidung* (Clothes) and *Wohnung* (house).

The test contains two sub themes, namely *Essen und Trinken* (food and drink) and *Kleidung* (clothes). Then, the researcher collected the materials for the test, e.g. texts and pictures. Moreover, the test blueprint is developed. There are some specifications of the test in the blueprint, namely the amount of the items (80 items), tested scope or competence, cognitive level, item number and the key answer.

C. Demonstrate

In this stage, the researcher developed the test items. The researcher wrote the item-card. After the item-cards have been written, the researcher tested the test's validity. Validity means that the test measures, what it tends to measure [11]. There are some kinds of validity, but every test should have good content validity. In order to achieve this, the validity of the test has been tested. This process has been done with expert judgment. The validators were four experts in different area, namely *Deutsch als Fremdsprache* (German as Foreign Language), German literature and evaluation.

There are some steps in the validation process. The steps are checking the appropriateness of theories and test blueprint, checking the appropriateness of test blueprint and items, scoring the items in the item-cards, whether the items are relevant or not, checking the key answer and scoring rubric and filling the questionnaire. The experts scored the questionnaire with values 1

until 4. 1 means not relevant, 2 means relevant enough, 3 means relevant, and 4 means very relevant. The data were analyzed with Aiken's V formula [12].

Based on the validity test, 68 items are relevant, 5 items should be corrected and 7 items are not relevant. The items, which are not relevant, can't be used. After the revision, the trial has been held.

The trial has been held in a school. The trial's goal is to know the readability of the test. The readability consists of right orthography, stem, instruction and pictures. The items for trial are 73 items. These are the relevant items of the validity test.

The respondents should not only answer the question but also give comments about the readability of the test. The test revision has been done based on the comments or questions from the respondents. The respondent have done the test and given comments.

A respondent could answer 65 items correctly. That is almost 90% correct. The lowest result is 41 items or 56 % can be answered correctly. The respondents needed different time to do the test. The test has been done between 50 minutes until 80 minutes.

Based on the trial result, it can be concluded that test has good readability. Based on this result, some revisions needed to be done. Then the second trial has been planned.

D. Develop

In this stage, the researcher prepared the second trial and checked the items. This stage lasted as long as there are some things that need to be corrected. The items are corrected based on the first-trial's result. After the revisions are made, the second trial would be done.

E. Deliver

In this stage, the second trial has been done. The respondents were 33 students of 11th grade class of language program. The second trial's goal is to know the item difficulty, item discrimination, reliability and distractor function. The item analysis was needed in order to get this information. The bad items have been cut and good items only are used.

Based on the item difficulty and item discrimination of an item, we can know, whether the item can be used or not. Item difficulty means how difficult or easy an item is for the test takers [7]. Beside item difficulty, item discrimination is important too. Item discrimination shows the ability of an item to differentiate the students' performance.

Not only item difficulty and item discrimination, but also the efficiency of distractors should be analyzed. Distractors are the false options of a multiple choice item [13]. Multiple choice item consist of a direct question or incomplete statement (stem) followed by two or more possible answer (called response), only one of which is to be selected [14]. The students should choose one correct answer, because in multiple choice item, there are some alternative answers, but there is only one correct answer, the other answers are distractors.

The distractors should be good. It means that a distractor of an item has been chosen by at least one test taker. The distractor, which is not effective, would be replaced. The items in the developed test have some distractors that should be replaced. The last aspect is the reliability. Reliability is the steadiness of a measure result [11]. Validity and reliability of a test is really important [6]. The reliability of this test has been tested with Alpha Cronbach. The reliability coefficient of the test is 0.831. It is considered as very high. After the item analysis, the amount of the good items is 54 items.

III. DEVELOPMENT RESULT

The development result is a test instrument for German language subject based on the curriculum 2013 for the 11th grade class language program in second semester. The test is intended to test the students' reading skill. The developed test is an objective test. The items are multiple choice items. There are five answer-choices, four as distractors and one as the correct answer. The test consists of some texts.

The test consists of eleven texts. These texts have different themes, eight texts contain *Essen und Trinken* (food and drinks) and three texts contain *Kleidung* (clothes) theme. Furthermore, the texts have different word amount. The short texts consist of 20 until 47 words and the students need 30 until 45 seconds to understand each text. Besides, the moderate texts consist of 57 until 82 texts. The students need a minutes to understand each text. The long texts consist of 114

until 126 words. The students need two minutes to understand each text. The table 1 below describes the texts in the instrument and followed by the example of the text and items.


Table 1. Texts in the instrument


Text	Items	Word amount	Time
1	4	114	2 Minutes
2	2	129	2 Minutes
3	4	78	1 Minute
4	4	23	30 Seconds
5	3	65	1 Minute
6	4	57	1 Minute
7	3	20	30 Seconds
8	3	82	1 Minute
9	3	47	45 Seconds
10	2	42	45 Seconds
11	4	63	1 Minute
12	2	24	30 Seconds
13	3	46	1 Minute


The example of short text with 20 words:


Peter findet elegante Kleidung auch am schönsten. Er trägt oft einen schwarzen Anzug, ein weißes Hemd, und eine rote Krawatte.


- Das Thema des Textes ist
 - Kleidungsfarbe
 - Kleidungsstücke
 - Kleidungsteile
 - Kleidungseinkaufen
 - Lieblingskleidung
- Peter trägt gern Krawatte.
Krawatte ist das Bild Nummer


a.


b.


c.



d.


e.
- Peter trägt oft einen schwarzen Anzug.
Das Gegenteil von schwarz ist
 - blau
 - weiß
 - rot
 - gelb
 - braun

Meanwhile, the example of moderate text with 78 words:

Haben Sie das gewusst? Deutschland steht im Safttrinken international auf Platz eins! Im Jahr 2004 hat jeder Deutsche circa 40 Liter Saft getrunken. Apfelsaft ist mit etwa 13 Litern im Jahr das beliebteste Saftgetränk. Viele finden, er schmeckt besser als Orangensaft.

In Deutschland mischt man Apfelsaft oft mit Mineralwasser. Das Getränk heißt Apfelschorle und schmeckt nicht so süß. Apfelschorle ist nicht nur bei Sportler ein Fitnessgetränk. Bei langen Autofahrten ist eine Apfelschorle in der Pause besser als Kaffee.

- Im Text geht es um
 - Fitnessgetränk
 - Getränk
 - Apfelschorle
 - Saft
 - Apfelsaft
 - Apfelsaft
 - mögen nur wenige Leute
 - schmeckt den Leuten nicht so gut
 - trinken die Autofahrer nicht
 - ist eine Mischung
 - trinken die meisten Leute im Jahr 2004
 - Diese Aussage sind **richtig**, außer:
 - Apfelschorle ist eine Mischung und schmeckt sehr süß
 - Mineralwasser ohne Apfel- oder Orangensaft ist gut für die Sportler
 - Bei langen Autofahrten sollte man viel Apfelschorle trinken
 - Nur wenige Leute trinken gern Orangensaft im Vergleich mit Apfelsaft
 - In Deutschland ist viel Saft von den Deutschen jedes Jahr getrunken
- 

The example of long text with 114 words:

Gaststätte * Hotelpension
 “ ZUR ERHOLUNG “
 Inh. 1. Stahl * 25551 Lockstedt * Tel. 04877/202

Wildgerichte

1. **Holsteiner Damhirschbraten**
 mit Champignons in Rotrahmsauce
 Preiselbeerbirne, Apfelrotkohl
 Salzkartoffeln 12,90 €

Rehbret aus umliegenden Jagdrevieren

2. **Zarter Rehkeulenbraten**
 Garniert mit frischen Champignons
 Rotweinsauce, Kiwi mit Preiselbeeren
 Prinzeßbohnen und Krokotten 14,80 €

Der immer beliebte

3. **Wildschweinbraten**
 in einer leichten Bordeauxsauce
 umlegt mit frischen Champignons
 dazu Apfelrotkohl und Krokotten 14,30 €
4. **Pikantes Wildpfännchen**
 Mariniertes Reh-Ragout
 mit Pfifferlingrahmsauce
 mit Preiselbeerbirne und Butterspätzle
 Apfelrotkohl 11,80 €

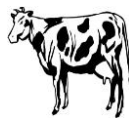
Zarter geht es nicht mehr

5. **Hirschrückenmedallions**
 In der Pfanne serviert
 Champignons und Sauce Hollandaise
 Broccolirößchen, Krokotten 15,50 €

Ein lohendes Ziel zwischen Hamburg und Kiel

7. Man findet den Text
- | | | |
|------------------|------------------|-------------|
| a. im Hotel | c. auf dem Markt | e. am Kiosk |
| b. im Restaurant | d. im Geschäft | |
8. Erika isst mariniertes Reh-Ragout. Sie muss Euro bezahlen.
- | | | |
|------------|------------|------------|
| a. 12,90 € | c. 14,30 € | e. 15,50 € |
| b. 14,80 € | d. 11,80 € | |
9. Sania kauft Wildschweinbraten, denn sie isst gern Schwein.
Schwein ist das Bild Nummer

A.



B.



C.



D.



E.



After all items were already written, the test validity has been tested. The developed test has good content validity. In testing the test validity, there are two steps that have been done, namely expert judgment and measuring with the Aiken's V formula. The validity coefficient is 0.71. It means that the test is valid. The criteria of validity are described in the table below [15].

Table 2. Validity Criteria

No.	Range	Category
1	$0,80 < r \leq 1,00$	very high
2	$0,60 < r \leq 0,80$	high
3	$0,40 < r \leq 0,60$	moderate
4	$0,20 < r \leq 0,40$	low
5	$-1,00 \leq r \leq 0,20$	very low

Moreover, the validity of the items is measured too. The amount of the items is 80 items. These items are intended to measure students' reading skill. The experts have checked the items and given some advices in order to make the items better and more relevant. The advices are: (1) false words should be avoided, (2) the indicators should be appropriate with basic competence, (3) the answer in the key answers should be same with the blueprint, (4) false written word should be corrected, (5) the texts sources should be written, (6) structure and vocabularies should be corrected, (7) using symbols in dialog should be avoided, (8) dialogs should have contexts, (9) complex sentences should be simplified, and (10) the pictures should be clear.

Then two trials have been done too. The first trial was intended to know the readability of the test. The time, which the respondents needed to answer the question, were noted. A respondent has done the test in 50 Minutes. He needed 13 minutes to read and understand the texts. He needed 30 seconds to answer each items. That was the fastest. A respondent has done the test in 80 minutes, 13 minutes for reading the texts, so he needed average time 55 seconds to answer each items. In conclusion, an item can be answered between 30 and 55 seconds.

The second trial was intended to know the quality of the items. The item difficulty, item discrimination, function of distractor, and reliability of the test were analyzed. Based on the item analysis result, the low-quality items were cut. The test has appropriate item difficulty and item discrimination. Besides that, the test also has good reliability. The reliability of the test has been measured with Alpha Cronbach formula. The reliability coefficient is 0,863. In conclusion, the test has good quality.

IV. CONCLUSION

The developed test instrument is a test for German language subject based on the curriculum 2013 for the 11th grade class language program in second semester. The test contains two sub themes, namely *Essen und Trinken* (food and drink) and *Kleidung* (clothes). The test is intended to measure students' reading skill. The items are developed based on the basic competence 3.3. In this basic competence, the students should understand text structure and cultural elements with *Alltagsleben* theme. The test items are multiple choice items.

The test validity and item validity have been tested with expert judgment. Moreover, two trials have been done to know the test readability and the quality of the items. The amount of the good items is 54 items. These items have good item difficulty, item discrimination, distractors and reliability.

The developed test has it strengths. Those are: (1) the test is appropriate with curriculum 2013, (2) the test tests the cognitive level from C1 until C4, (3) the test has good items, (4) the test has good validity and reliability, (5) the test measures students' reading skill, (6) teachers are able to use the items of this test.

ACKNOWLEDGMENT

The researchers thank to all person that have been involved in this research.

REFERENCES

- [1] Kementerian Pendidikan dan Kebudayaan Indonesia, Undang – Undang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional. Jakarta: Kemendikbud, 2003.
- [2] H. E. Mulyasa, Pengembangan dan Implementasi Kurikulum 2013. Bandung: Remaja Rosdakrya, pp.60.
- [3] Kementerian Pendidikan dan Kebudayaan Indonesia, Kurikulum 2013 Kompetensi. Jakarta: Permendikbud, pp. 183.
- [4] B. Nurgiyantoro, Penilaian dalam Pengajaran Bahasa dan Sastra. Yogyakarta: BPFE, pp. 5.

- [5] P. P. Lestari, Validitas dan Reliabilitas Soal UAS Bahasa Jerman Kelas XI IPS SMA Negeri 7 Malang. Unpublished.
- [6] S. Arikunto, Dasar – Dasar Evaluasi Pendidikan (Revisi Kedua). Jakarta: Rineka Cipta, 2012.
- [7] B. Nurgiyantoro, Penilaian Pembelajaran Bahasa Berbasis Kompetensi. Yogyakarta: FEB UGM, 2013, pp. 108.
- [8] K. Cennamo & D. Kalk, Real World Instructional Design. Canada: Thomson Learning, Inc, 225, pp 5-7.
- [9] N. S. Sukmadinata, Metode Penelitian Pendidikan. Bandung: Remaja Rosdakarya, 2009, pp. 164.
- [10] Sugiyono, Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung: Alfabeta, 2010, pp. 147.
- [11] H-G. Albers & S. Bolton, Tetsen und prüfen in der Grundstufe. München: Langenscheidt, 1995, pp. 22-25.
- [12] S. Azwar, Penyusunan Skala Psikologi (Edisi 2). Yogyakarta: Pustaka Pelajar, 2012, pp. 113.
- [13] ALTE. Handbuch zur Entwicklung und Durchführung von Sprachtests, Zur Verwendung mit dem GER Frankfurt am Main: telc GmbH, 2012, pp.17.
- [14] D. W. Johnson & R. T. Johnson, Meaningful Assessment (A Manageable Cooperative Process). Boston: A Pearson Education Company, 2002, pp. 62.
- [15] J. P. Guilford, Fundamental Statistics in Psychology and Education. New York: Mc. Graw-Hill Book Co. Inc, 1956, pp. 145.

EVALUATION OF MANAGERIAL LEADERSHIP ABILITY OF SENIOR HIGH SCHOOL HEADMASTERS IN SLEMAN

Sabar Budi Raharjo, Lia Yuliana
(Puslitjak Balitbang Kemdikbud and FIP UNY)
e-mail: raharjo2sbr@yahoo.co.id
e-mail: yulianauny@gmail.com

Abstract - The aim of this study is to find out the ability of managerial leadership of SMA's headmasters at Sleman Regency, Special District of Yogyakarta. The method of this study uses evaluation with strategic survey. The sample of the Headmasters is taken from 4 different SMAs which are considered representing the good, fair and poor categories. Those 4 SMAs are SMA N 2 Sleman, SMA N 1 Pakem, SMA N 1 Depok and SMA N 1 Kalasan. Therefore, the subject of this research is 4 headmasters and 16 teachers which are collected randomly from the four schools. The data collection uses questionnaire technique to assess the managerial ability of the headmasters. The data analysis technique uses descriptive data analytical technique along with evaluation criteria. The result of the study shows that the ability of managerial leadership of the headmasters according to the assessment of 4 headmasters is 1.94 and falls in a very good category. The managerial ability of the headmasters which is assessed by 16 teachers is 1.90 and falls in a good category. Thus, the result of this study shows that the managerial leadership ability of the headmasters at Sleman regency is good according to the self-assessment of the headmasters. The teacher's assessment shows a good result as well.

Keyword: *ability, managerial, headmaster*

I. INTRODUCTION

The research result of Sudarwan stated that one characteristic of education in Indonesia is still not able to show an adequate of work performance. One of the reasons is the lack of the headmaster's role in improving teacher performance due to the lack of effective leadership. The headmasters pay less attention to the progress of teacher performance because of the ability of the learning model that does not take notice on substantive aspects. This phenomenon cannot be denied given the need in reforming national education system involving various components systematically (Sudarwan, 2002).

Headmasters' leadership can be an institution machine for the teachers and school's civitas to improve their work culture. The organization's success in achieving the goals that have been set will be highly dependent on the critical role of the leaders (Dufour & Barkey, 2005). Headmasters have leadership ability to influence all parties involved in educational activities in schools, which is to cooperate in implementing the objectives of the schools (Agus Supriyono, 2012).

Therefore, to improve the school performance and work culture, it can be done by increasing the ability of headmasters' leadership in terms of learning, organization and managerial (Diwijaya, 2009). School culture is one of the factors that influence the work culture. With a conducive work culture in the school, the students will achieve satisfactory academic achievement. This good academic performance is what ultimately impact on school improvement optimally. Therefore, the headmasters' leadership is very important for teachers and education personnel to improve their performance (Kumaris, V. 2014).

School leadership is one of the inputs for schools in performing their duties and functions as well as the effect on the process of schooling (Slamet PH, 2001).

Holified & Cline (2007) state that one of the main duties of headmasters is to improve the teachers' performance. This is in line with the statement of Dufour & Barkey (2005) that a success in improving the school depend on the development of profession within the school, and the essential one is the professional development of teachers.

As a leader, a headmaster conducts new approaches in order to improve the school capacity (Husaini Usman, 2010). According to Harold J. Levitt, it is stated that manager or leaders, in one way or another, must influence other people to do what managers want them to do (2005). It can be seen that headmasters as managers in school hold the most important role to achieve the school objectives that has been written in the vision and mission of the schools.

This study focuses on evaluation of headmasters' managerial ability in Senior High Schools. The problems of this study is how good the managerial leadership ability of Senior High Schools' Headmasters. Therefore, the objective of this study is to find out the managerial leadership ability of Senior High Schools' headmasters in Sleman, Special District of Yogyakarta

II. RESEARCH METHOD

The method used is the evaluation of research or survey research evaluation strategy. The sample of this study is 4 Senior High School headmasters that are considered to represent the school category of high, medium, and low. The four schools are SMA N 2 Sleman, SMA N I Pakem, SMA N I Depok and SMA N I Kalasan. Therefore, the subject of this research is 4 headmasters and 16 teachers which are collected by proportional random sampling from the four schools. The data collection uses questionnaire technique to assess the managerial ability of the headmasters. The indicators of managerial ability which are evaluated are the ability in terms of conducting regular meeting regarding school management, providing chances for teachers and vice principal to submit proposals for the development of the school, defining target date for completion of the school program, conducting scrutiny of the programs of the school development, involving the vice principal in preparing the target date for completion of the school program, involving teachers in preparing a target date for completion of the school program, involving educators in preparing a target date for completion of the school program, completing tasks and work in accordance with a predetermined time, doing observation activities in a planned management, developing an evaluation system of school programs, raise the school program evaluation team, raise the monitoring team of the implementation of the school program, conducting scrutiny of the evaluation results, conducting reflection on the results of the implementation of the program, carrying out the follow-up results of the evaluation program, and set up new programs that are considered necessary and qualified. The data analysis techniques used is the descriptive data analysis using evaluation criteria analysis of Nana Sudjana, which is 1-5 converted into 0-2 scale. The criteria are: > 1.8 = very good, $> 1.4 - 1.8$ = good, $> 1.0 - 1.4$ = fair, $> 0.6 - 1.0$ = poor, and < 0.6 = very poor (Nana Sudjana, 2005).

III. RESULTS AND DISCUSSION

In the education activities in the level of school, headmasters are leaders who bring the whole school civitas to the success or failure in the education implementation. In the level of classroom, teachers are the ones who responsible for the improvement of the students' learning. The success of the students' learning is highly affected by the quality of the teachers in the learning process, while the quality of the teachers is also affected by the leadership of the headmasters in operating the school management and the learning process. The better the teacher's performance, the better the students' achievement in learning. The better the headmasters' leadership in operating the school management and learning, the better the teachers' performance in doing the learning process. As a result, the school quality will get better.

This study focuses on the managerial leadership ability of the Senior High Schools Headmasters in Sleman. The sample of this study is SMA N 2 Sleman, SMAN I Pakem, SMA N I Depok, dan SMA N I Kalasan. Therefore, the subject of this study is 4 headmasters and 16 teachers which were taken by proporsional random sampling from the four sample schools. The selection of the four schools was conducted by proporsional sampling that is schools with the level of quality is almost the same but have a clear rank based on the the result of national examination obtained by SMA N 2 Sleman, SMA N 1 Pakem, SMA N 1 Depok, and SMA N 1

Kalasan. The evaluation of the headmasters' managerial ability was conducted on June 2015. The selection in June is intended at least to the reference of development and improvement in the implementation of the principal tasks in the next year.

Table 1. The Result of Headmasters' Self-Assessment towards Managerial Leadership Ability

No	Indicators of Managerial Ability	Score
1	Conducting regular meeting regarding school management	2.0
2	Providing opportunities to the teachers and vice principles to propose ideas of school development	2.0
3	Set a deadline in completion of the school programs	2.0
4	Doing scrutiny towards school development programs	2.0
5	Involving the vice principles in preparing deadline for completion of the school program	2.0
6	Involving teachers in preparing deadline for completion of the school programs	2.0
7	Involving educators in preparing deadline for completion of the school programs	2.0
8	Completing tasks and works with a predetermined time	2.0
9	Doing a planned management activity observation	2.0
10	Developing school program evaluation system	1.75
11	Promoting a school program evaluation team	1.75
12	Promoting a monitoring team of school program implementation	1.75
13	Evaluating the programs together with the evaluation team	1.75
14	Doing scrutiny towards the evaluation results	2.0
15	Reflecting on the program implementation results	2.0
16	Carrying out the follow-up of evaluation program results	2.0
17	Preparing new programs that are considered necessary and qualified	2.0
	Total Score	33
	Mean	1.94

Based on the headmasters' self-assessment towards managerial leadership ability shows that the total score is 33 and the mean of all items is 1.94 which means that the managerial leadership ability of the senior high schools' headmasters in Sleman is very good. Each item of 1-9 obtained score 2 which means very good, items of 10-13 obtained score 1.7 each which means good, and items of 14-17 obtained score 2.0 each which means very good. The total score and the mean shows a very good result. The validation of the headmasters' self-assessment result towards managerial leadership ability is conducted by doing assessment to the teachers every day and has contact with the headmasters directly. The result is almost the same as it shown in the following table.

Table 2. The Result of Teachers' Assessment towards Managerial Leadership Ability of the Headmasters

No	Indicators of Managerial Ability	Score
1	Conducting regular meeting regarding school management	2.0
2	Providing opportunities to the teachers and vice principles to propose ideas of school development	2.0
3	Set a deadline in completion of the school programs	2.0
4	Doing scrutiny towards school development programs	2.0
5	Involving the vice principles in preparing deadline for completion of the school program	2.0
6	Involving teachers in preparing deadline for completion of the school programs	1.82
7	Involving educators in preparing deadline for completion of the school programs	1.76
8	Completing tasks and works with a predetermined time	2.0
9	Doing a planned management activity observation	2.0
10	Developing school program evaluation system	1.64
11	Promoting a school program evaluation team	1.64
12	Promoting a monitoring team of school program implementation	1.75
13	Evaluating the programs together with the evaluation team	1.75
14	Doing scrutiny towards the evaluation results	2.0
15	Reflecting on the program implementation results	2.0
16	Carrying out the follow-up of evaluation program results	2.0
17	Preparing new programs that are considered necessary and qualified	2.0
	Total Score	32.36
	Mean	1.90

Based on the 16 teachers' assessment result which each school was represented by 4 teachers, it shows that the managerial leadership ability obtained the total score of 32.3 and the mean of all items is 1.90. Based on the criteria used, it means that the managerial leadership ability of the senior high schools' headmasters in Sleman is very good. Each of 1-5 items obtained score 2 which means very good, item obtained score 1.82 which means very good, item 7 obtained score 1.7 which means good, Each item of 10-11 obtained score 1.64 which means good, Each item of 12-13 obtained score 1.75 which means good, and each item of 14-17 obtained score 2.0 which means very good. The total score and the mean of the managerial leadership ability of the headmasters fall in a very good category.

If it is wholly-evaluted based on headmasters' self-assessment and teachers' assessment toward managerial ability if the headmasters, the total score obtained is 65.36 and the mean is 1.92. It shows that the headmasters' managerial ability based on those two groups of headmasters and teachers' evaluator is very good. Thus, it can be justified that the managerial leadership ability of senior high schools' headmasters in Sleman is very good. This becomes a basic modal for the improvement of the school quality in order to compete in the national level.

In line with Robinson study (2009) that conducted meta-analysis towards 12 out of 13 researches about headmasters' leadership and the results showed that the effect size of headmasters' leadership had a significant effect towards the improvement of learning quality which resulted on the whole school quality. The effect size of headmasters' leadership towards teachers' performance and learning quality is three times wider than the effect size of transformational leadership. The findings showed that the headmasters' leadership has great influence for the success of the educational programs thorough the improvement of the functional quality and improvement of learning quality systematically.

According to Harold J. Leavitt (2005), manager or leaders, in one way or another, must influence other people to do what managers want them to do. A leader must have the ability to influence others to do what he or she wants. It means that a headmaster as a manager in school holds the most essential role to achieve the school's objectives that have been written in the vision and mission of the school. In this case, a headmaster who is a leader in school has main tasks in managing the implementation of the school activities. In operational terms, the main duties of the headmaster include empowering the whole resources in order to achieve school's objectives effectively and efficiently.

The result of Herlinger study (2009) which conducted review toward 40 research results about the effect of headmasters' leadership towards school quality showed that the indirect effect was more consistent than the direct one. It means that the headmasters need to concentrate on managerial leadership. The efforts of headmasters to improve students' achievement cannot be done directly but through improvement of teachers' performance. The headmasters must have knowledge about learning theory, instructional and curriculum. Headmasters' managerial leadership is directly related to the improvement of school program quality including programs which are school achievement-oriented.

IV. CONCLUSION

Based on the results and discussion, it can be concluded that as follows. The managerial leadership of the headmasters based on the assessment of 4 headmasters is 1.94 which falls in a very good category. While the managerial ability of the headmasters which was evaluated by the 16 senior high school teachers is 1.90 which falls in a good category. The results are wholly evaluated based on headmasters' self-assessment and teachers' assessment toward managerial ability of the headmasters, the total score obtained is 65.36 and the mean is 1.92. This shows that the managerial ability of the headmasters, based on the two groups that are headmasters and teachers evaluator, is very good. Thus, it can be justified that the managerial leadership ability of the senior high school headmasters in Sleman is very good. This very good result is a basic modal for the school quality improvement of both academic field and other related fields.

V. SUGGESTIONS

Based on the result of this study, there some suggestions formulated as follows.

1. Giving suggestions to the school supervisors to utilize the result of this study in evaluating managerial ability of the senior high school headmasters to improve the functional quality of the school both in academic field and non-academic field.
2. Suggestions are given to the headmasters based on the result of managerial ability assessment obtained, to always improve their ability especially in managerial field.
3. For the future researcher, the result of this research focuses on managerial ability area and still can be developed with the wider research area and the variety of headmasters' leadership ability such as learning leadership and organizational leadership.

REFERENCES

- [1] Agus Supriyono. (2012). Membentuk Kultur Pembelajaran yang Mendidik. *Jurnal Paramita*, Vol 22 Nomor 2 hlm. 129-140.
- [2] Dufour, R & Barkey, T. (2005). *The Principals as Staff Developer*. Journal of Development, Fall 2005 (Volume 16. Nomor 4).
- [3] Diwijaya. (2009). *Kompetensi Guru: Citra Guru Profesional*. Bandung: Alfabeta.
- [4] Herlinger & Heck. (2009). *Exploring the Principal Contribution to School Effectiveness*. Boston: Pearson Education.
- [5] Holifield, M & Cline, D. (2007). "Clinical Supervision and its outcome: teacher and Principals Report", *National Association of Secondary School Principals, NASSP Bulletin; Research Library*.
- [6] Husaini Usman. (2010). *Manajemen, Teori, Praktik, dan Riset Pendidikan*. Jakarta: Bumi Aksara.
- [7] Kumaris, V. (2014). Soft skills and teacher accountability in the context of quality education. *Journal of Education and Psychological Research*, 3, 79-80.
- [8] Leavitt, Harold J. (2005). *Top Down: why hierarchies are here to stay and how to manage them more effectively*. Boston, Massachusetts: Harvard Business School Publishing, 60 Harvard Way.
- [9] Robinson, V.J.M. (2009). *School Leadership and student outcome: Identifying What Work and Why: Best Evidence Synthesis Iteration*, New Zealand: Ministry of education.
- [10] Slamet PH. (2001). "Karakteristik Kepala Sekolah Tangguh." *Jurnal Pendidikan dan Kebudayaan*. (No. 025 tahun VI). Halaman 319-333.
- [11] Sudarwan. (2002). *Inovasi Pendidikan dalam upaya Peningkatan Profesionalisme Tenaga Kependidikan*. Bandung: Pustaka Setia.

EVALUATION OF SOCIAL ATTITUDE CORE COMPETENCE (KI-2) IMPLEMENTATION IN STATE ELEMENTARY SCHOOL IN YOGYAKARTA

Siti Aminah¹⁾, Yulian Sari²⁾

¹ Graduate School Program of Yogyakarta State University

² Graduate School Program of Yogyakarta State University

¹nceekaminah@gmail.com, ²yulialahwan@gmail.com

Abstract - The purpose of this research is to describe the implementation of the social attitude core competence (KI-2) in state elementary schools in Yogyakarta in terms of: understanding and planning. This research is an evaluation program with the quantitative descriptive approach, and the evaluation model used is the Discrepancy Model. The subjects in the evaluation were teachers and principals. The sample was established using the purposive random sampling technique. The variables in this study included social attitude core competence (KI-2). The techniques and instruments used in this evaluation were an observation by using observation sheets, study of lesson plans using worksheet, The content validity using Aiken's formulation was used as the instrument validity. The instruments' reliability was obtained by applying Cohens' Kappa technique calculation. The results of this research are as follows. First, the implementation of spiritual attitude core competence in state elementary schools in Yogyakarta City is in a obtains the score of 19.17 and is categorized as not good. Second, the planning conducted by the teachers in state elementary schools in Yogyakarta City regarding the integration of social attitude core competence in lesson plan is categorized as very good with the score of 36.

Keywords: *Evaluation, Program Implementation, Social Attitude Core Competence (KI-2).*

I. INTRODUCTION

Indonesian government takes some actions in order to increase the quality of Indonesian education. One of the government efforts is creating a curriculum as the reference in conducting education program. A curriculum is a set of plan that is applied in a teaching learning process. The government also develops curriculum 2013 as another effort to improve the quality national education program [1].

The aim of curriculum changing is to balance the cognitive, affective, and psychomotor domain of the learners. In accordance with that, Abdul Majid argues that developing of curriculum 2013 is oriented to improve and balance the learners' attitude competence, skill, and knowledge. Therefore, the aim of the changing and the developing of curriculum 2013 is to balance the cognitive, affective, and psychomotor domain of the learners [2].

One of the important points in the implementation of curriculum 2013 is the core competence (KI) [3]. Core competence (KI) refers to the realization of SKL in the form of the qualification that the learners who have finished particular education degree should have. It can be seen from the main competence that is categorized into affective, cognitive and psychomotor domain that should be learned by the learners in a degree, class, and subject [4]. It can be concluded that core competence refers to qualification that should be obtained by a learner from the teaching learning process. According to PP number 67 [4]. there are four core competences that are related each other as follows:

Core competence -1 (KI-1): spiritual attitude core competence

Core competence -2 (KI-2): social attitude core competence
 Core competence -3 (KI-3): knowledge core competence; and
 Core competence -4 (KI-4): skill core competence

It is expected that the teacher should teach the four core competences in every subject. Social attitude core competence is one of the important competences among the four core competences that should be obtained by the learners. Social attitude competence refers to the value related to relation between human being with God, with another human being, with animals, or things. It is expected that social attitude competence is able to make the learners can interact with another human being and their environment [5].

The aim of KI-2 implementation in Curriculum 2013 is the learners are expected to have good attitude and good interaction with the other people and their environment [6]. It indicates that it is important to implement KI-2 appropriately. It is expected that the learners can achieve the learning goals not only in cognitive domain but also in affective and psychomotor domain. In addition, through the implementation of KI-2 in teaching learning process, it is expected that the students will have good attitude through the integration of KI-2 in the teaching learning process.

However, based on the pre-survey and initial interview with the principals and class teachers of fifth grade in state elementary schools in Yogyakarta, generally, the teachers encounter some difficulties in implementing the KI-2. One of the difficulties that the teachers encounter is it is hard to integrate the learning material with social attitude core competence. One of the factors that drives this problem is a lack of teacher' knowledge and reference.

Another difficulty is it is hard to implement the KI-2. It is because knowledge core competence (KI-3) and skill core competence (KI-4) can easily explicitly seen. It means that the knowledge core competence (KI-3) and skill core competence (KI-4) is clearly stated in the learning activities in the students' text book. On contrary, social attitude competence is an implicit object. It means that it is not explicitly stated in students' text book, so it should be taught indirectly (indirect teaching) by the teachers when they deliver the KI-3 and KI-4 in teaching learning process. Besides, the teacher competence also influences the delivering KI-2 in teaching learning process because the teachers will have their own method in delivering the KI-2 in teaching learning process.

Furthermore, another hindering factor in KI-2 implementation in teaching learning process is implementing KI-2 is something new for them, so they are not used to including it in teaching learning process. The teachers usually prefer to prior delivering the KI-3, the knowledge competence, so most of the teachers usually deliver more learning material to achieve the knowledge competence and forget about the KI-2. Those difficulties cause the teacher cannot optimally deliver the KI-2 in teaching learning process. Those difficulties also give impact on the students. KI-2 that is not delivered optimally causes the students cannot gain the competence goal optimally.

The implementation of KI-2 should be involved in the whole process of teaching and learning, from the beginning until the end of the teaching learning process. As a result, the value of KI-2 can be implemented in earlier degree such as elementary school because elementary school level has more portion than the higher education level. Besides, based on the initial interview with the teachers of state elementary schools in Yogyakarta that employ curriculum 2013, so far evaluation of the implementation of the curriculum has not been conducted, especially the evaluation of KI-2 implementation of curriculum 2013 in those schools.

Regarding the explanations above, it is important to conduct an evaluation research. The study entitled "the Implementation of Social Attitude Competence of Curriculum 2013 in State Elementary Schools in Yogyakarta" is important to be conducted. The study attempts to seek the planning and the implementing of KI-2 that is conducted by the teachers in teaching learning process. As a result, it is expected that the KI-2 can be implemented optimally by the educators in teaching learning process, and the objective of curriculum 2013 can be achieved optimally.

II. METHOD

The type of evaluation in this study is program evaluation because it aims to evaluate the implementation of social attitude core competence (KI-2) in state elementary schools in Yogyakarta. The evaluation is conducted in fifth grade of elementary schools in Yogyakarta with the teachers as the subject. The approach used in this study is quantitative descriptive approach. The evaluation model that is adapted in this study is Discrepancy evaluation model that is

postulated by Malcom Provous [7]. This study adapts this model because the researcher attempts to identify whether there is any discrepancy or not between what should happen based on the standard (S) of the implementation and the performance (P) of the implementation in the field regarding KI-2. *PP nomor 67* becomes reference of KI-2 implementation standard (s) that will be the standard of comparison to evaluate the performance (P) of the evaluation object. Discrepancy (D), as the result of comparing between (S) and (P) facilitates the researcher to identify the implementation of social attitude core competence (KI-2) in state elementary schools in Yogyakarta.

The research was conducted in state elementary schools in Yogyakarta that implement curriculum 2013. The subject of this research was all of the teachers of fifth grade in state elementary schools in Yogyakarta that implement curriculum 2013. The data were obtained from a teacher of fifth grade in every school, so there were four teachers who became the object of this research. The object of this research was four schools that were obtained by purposive sampling technique.

The collecting data technique in this study was observation, and documentary study. The main collecting data technique in this study was the observation using observation protocol that contained of indicators regarding the observed attitudes. The documentary study was conducted to analyze the lesson plan (RPP).

The validity used in this evaluation study was content validity. The content validity of the instrument was obtained from the expert judgment. The instrument was given to the expert to be examined the validity of the instrument. Furthermore, it used aiken's formula to calculate the validity. Besides, the calculation of reliability coefficient was obtained by using Cohen's kappa formulation. The data would be calculated by using computer program, SPSS.

III. FINDINGS AND DISCUSSIONS

The findings of social attitude core competence (KI-2) implementation in state elementary school in Yogyakarta obtains the overall score 19.17, so it can be categorized as poor. The result was obtained by comparing the obtained score with table of evaluation success criteria. The evaluation success criteria of social attitude competence (KI-2) implementation in state elementary schools in Yogyakarta can be seen in the following table.

Table 1. Table of Categorization for Evaluation of Social Attitude Core Competence (KI-2) Implementation.

Criteria	Score
Very Good	$X \geq 30$
Good	$30 > X \geq 20$
Poor	$20 > X \geq 10$
Very Poor	$X < 10$

Based on the table of criteria above, it is clear that the result of social attitude core competence (KI-2) implementation in state elementary schools in Yogyakarta is as follows:

Table 2. The result of Evaluation of Social Attitude Core Competence (KI-2) in State Elementary School in Yogyakarta

The Implementation of KI-2	Score	Percentage	Categorized
Honesty	13,8	34,50	Poor
Discipline	17	42,50	Poor
Responsibility	25,8	64,50	Good
Courtesy	15,20	38,00	Poor
Care	17,8	44,50	Poor
Self-Confidence	31,3	78,25	Very good
Patriotism	13,3	33,25	Poor

The result of (KI-2) generally obtains score 19.17 with percentage 47.2. As a result, the implementation is categorized as poor. However, if it is examined by the detail of the result from each sub indicator of KI-2, the result is self-confidence is as sub indicator that obtains the highest score among the sub indicators of KI-2 with score 31.3 and percentage 78.25. As a result, the implementation is categorized as very good. On contrary, patriotism sub indicator obtains the lowest score with score 13.3 and percentage 33.25, so it is categorized as poor. Besides, another sub indicator that is categorized as poor is honesty that obtains score 13.8. Courtesy sub indicator obtains score 15.20. Discipline sub indicator obtains score 17.0. Furthermore, care

obtains score 17.8. The last indicator, responsibility obtains score 25.8, and it is categorized as good. In order to make it clearer, the detail of KI-2 implementation can be seen in the following graphic.

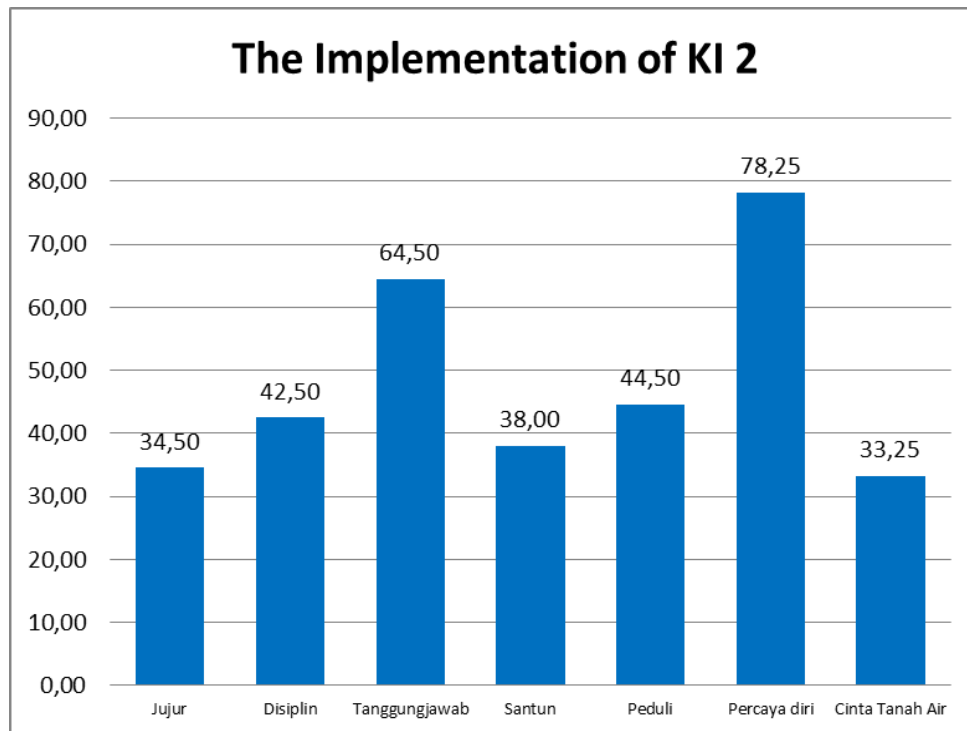


Figure 1. The Implementation of Social Attitude Core Competence (KI-2) in State Elementary School in Yogyakarta

Beside the implementation of KI-2 that is obtained by the observation to the teachers of state elementary schools in Yogyakarta, the implementation of the social attitude core competence is also examined by the planning that is conducted by the teachers. The planning can be seen in the integration of social attitude values in lesson plan (RPP). It means that the lesson plan that the teachers create will be examined by three experts to seek whether the values of KI-2 are integrated or not in the lesson plan by the teachers.

Based on the lesson plan analysis, it obtains that the planning that is conducted by the teachers in integrating the values of social attitude core competence into lesson plan can be categorized as very good with score 30 and percentage 100%. In another words, the lesson plan that is made by the teachers of state elementary school in Yogyakarta has already contained of social attitude values that will be delivered to the students in teaching learning process. The social attitude values that are stated in the lesson plan that is made by the teachers of state elementary school in Yogyakarta are as follows: (1) honesty; (2) discipline; (3) responsibility; (4) courtesy; (5) care; (6) self-confidence; (7) patriotism.

IV. CONCLUSIONS

According to the findings regarding the implementation of social attitude core competence (KI-2) in state elementary school in Yogyakarta, it can be concluded as follows:

1. The planning that is conducted by the teachers of state elementary school in Yogyakarta regarding the integration of social attitude core competence in the lesson plan is categorized as very good. It is proven by analyzing the lesson plan that is made by the teachers that obtains score 30. It means that the teachers have integrated the values of social attitude core competence into the lesson plan.
2. The implementation of social attitude core competence in state elementary school in Yogyakarta is categorized as poor. It is proven by the obtained score 19.17 with percentage 47.925.

REFERENCES

- [1] Majid & Chaerul Rochman. *Pendekatan ilmiah dalam implementasi kurikulum 2013*. Bandung: Remaja Rosdakarya. 2014.
- [2] Abdullah. *Pengembangan kurikulum teori & praktik*. Jakarta: Raja Grafindo. 2014.
- [3] Mulyasa. *Pengembangan implementasi kurikulum 2013*. Bandung: Remaja Rosda Karya. 2014
- [4] Kemendikbud.2013. *Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 67 Tahun 2013 Pasal 1 Tentang Kerangka Dasar Dan Struktur Kurikulum Sekolah Dasar/Madrasah Ibtidaiyah*.
- [5] Kemendikbud.2013. *Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 64, Tahun 2013, Tentang Standar Isi Pendidikan Dasar Dan Menengah*.
- [6] Tim Pengembangan MKDP Kurikulum dan Pembelajaran. *Kurikulum dan pembelajaran*. Jakarta: Rajawali Pers. 2011
- [7] Fitzpatrick, J. L., Sander, J. R., & Worthen, B. R. *Program evaluation: alternative approaches and practical guidelines*. Upper Saddle River: Pearson Education. 2011.
- [8] Dana Clark Baldwin. Spiritual Identity: Evaluating a Seminar on Spiritual Identity on Spiritual Wellness. *Journal by ProQuest LLC*. 2008

THE EVALUATION OF THE FOREIGN LANGUAGE INTENSIFICATION PROGRAM FOR THE STUDENTS OF UIN AILAUDDIN MAKASSAR

Sitti Mania

Alauddin State Islamic University in Makassar
sitti_mania@yahoo.com

Abstract -This study is aimed to evaluate the implementation of the program of intensification of foreign languages at UIN Alauddin Makassar with the following research questions, 1) to what extent the context of components are supporting the implementation of intensification foreign language program ?, 2) to what extent the component is supporting the implementation of intensification foreign language program ?, 3) to the extents of the components process is implemented on the intensification of foreign language program ?, and 4) to what extent the contribution of intensification foreign language program is improved for student's skill ?. Evaluative research using CIPP is a model that was implemented in 2014/2015 academic year, by using research instruments, such as questionnaires, interviews and observation guide. The sample was 3 of 8 Faculties were selected by purposively. Based on the research and the data analysis is concluded that the intensification of foreign language program is run well. Several Conclusions can be underlined from this study are: 1) the suitability of intensification of foreign languages for new students are viewed from component context is Appropriate, 2) the condition of the factors supporting the intensification program of foreign language seen from the component inputs are good, 3) the implementation of intensification for eign languages program is run well, 4) The foreign language program can be seen from the intensification of the component products are already succesful.

Keyword: *intensification foreign language program, UIN Alauddin Makassar, CIPP Program.*

I. INTRODUCTION

Alauddin State Islamic University in Makassar as one of the institutions of higher education based Islamic mission is to integrate the science of religion and science and technology. One of the efforts made to achieve the mission is to develop foreign language learning by establishing a technical executing unit named intensification foreign language program.

Program organized intensification of foreign language teaching and learning of English and Arabic. This program was born as a form of reflection of the function and purpose of national education set forth in Law No. 20 Year 2003 on National Education System states that education at all levels, including higher education should be organized in a more terkonsepsi, planned, and systematic, so as to compete, ethical, moral, politeness and interact with the community and in the wider global.

The intensification of foreign language programs have a function and authority for planning and program development language and the language of education and teaching in the form of a special learning program outside the regular classroom to all new students for two semesters.

Program intensification of foreign language basically has several times undergone a name change. At the beginning of its implementation in 1997 known as the matriculation program, students learn in classes according to their respective faculties and implementation of activities handled by kepanitian at institute level. In the next development changed its name to the intensification of foreign languages program management submitted to the faculty level. Along with the demands for students to have the competence to communicate in a foreign language, then in the academic year 2003/2004 was born a new program known as pious enlightenment programs and life skills (PIKIH). In the next phase, PIKIH changed its name, back to P rogram I

intensifikasi Bahasa Arab. At this time, the activities are not only focused on learning a foreign language, but also carries the mission of the character formation of students. The name change, strongly associated with the ideal expectations of the existence of the program.

Based on the background, then the research problem is formulated as follows: 1) the extent to which components of the context of providing support to the implementation of the program of intensification of foreign languages at UIN Alauddin Makassar ?, 2) the extent to which the input component to support the implementation of the program of intensification of foreign languages at UIN Alauddin Makassar ?, 3) the extent of adherence to process components on the intensification of foreign language program at UIN Alauddin Makassar ?, and 4) the extent to which the contribution program intensification of foreign languages at UIN Alauddin Makassar in improving the foreign language skills of students?.

This study aimed to examine the effectiveness of the implementation of the program of intensification of foreign languages in shaping foreign language skills of students. The results obtained through this research is expected to contribute positively to the implementation of a higher quality program.

II. RESEARCH METHODOLOGY

The research was carried out at UIN Alauddin Makassar in the academic year 2014/2015. The study focused on three faculties of 8 faculties using purposive sampling. The research method evaluative research with case study approach that aims to produce a detailed description of a phenomenon, develop explanations which can be given of the case study and evaluate the phenomenon. The nature of this evaluation was to see how effective the implementation of the program of intensification of foreign languages, especially Arabic language learning at UIN Alauddin Makassar. The model used is the CIPP evaluation model developed by the National Study Committee on Education of Phi Delta Kappa led by Stufflebeam and friends.

The instrument used in this study consists of: 1) a questionnaire, 2) interview, 3) guidelines for observation, and 4) document. Sources of data include: 1) a document in the form of written reports and other physical evidence, 2) Head of Language Center UIN Alauddin Makassar, 3) Chief Executive program, 4) faculty and staff, 5) students involved in the program activities, and 6) other personnel associated with the program. Data were analyzed by using domain analysis techniques, taxonomy, and componential based on the stages of evaluation. The analysis was conducted concurrently with data collection. Stabilization of the validity of the data obtained was done by using triangulation.

III. RESULTS AND DISCUSSION

A. *Component Context Intensification Program in Foreign Languages UIN Alauddin Makassar*

The research component focused on the context of efforts to uncover unmet needs of students and the learning environment in order to achieve the objectives of the program. The purpose of learning Arabic in the intensification of foreign language learning program differs from regular Arabic language learning in every faculty. At the faculty level, learning Arabic is targeting the four language skills of listening, reading, speech, and writing skills, while the learning Arabic program focuses only on the ability to speak.

Because the purpose of the program is the ability to speak, then material is material to hone the ability to speak in Arabic. Material and teaching material shaped into a handbook muqarrar and teachers and students during the learning process. Material conversation includes daily student activities both at home, in the dorm, in the library, cafeteria, and on campus.

On the other hand, there are some students who expect learning Arabic is not limited to the speaking skills, but also on the ability to read a variety of literature speak Arabic, speak Arabic writing scientific papers, study and understand the meaning of verses and hadiths armed with knowledge of the Arabic language.

To answer the expectations of these students, then manager designed several activities supporting language acquisition, such as: 1) students are required to speak Arabic in the dorms either with caregivers dormitories or with fellow students, 2) students are required to attend additional learning in the hostel as conducting studies, language games, watch a film about the language and so on, 3) end of each semester, students attend several competitions related to language, such as speech contest, quiz Arabic, Arabic writing essays, humorous stories Arabic and religious song contest.

One important thing that should be of concern is the program manager of the environment as well as a means of supporting learning activities. From the results of questionnaire analysis found that there are 65% of the students said that the learning environment is not conducive.

B. Component Input Intensification Program in Foreign Languages UIN Alauddin Makassar

In the academic year 2014/2015 the number of new students who take the program intensification of foreign languages as much as 5,263 students spread at 8 faculties. The number of students was so great that educational background is diverse, so the ability of their initial program intensification of foreign languages, especially in learning Arabic is quite varied. Among them are some who have already learned Arabic before so they understand when they hear the phrase in Arabic, has been can read and write in Arabic well. among them are some who never learned Arabic before, there are even among those who can not read and write in Arabic.

Associated with the condition tersebut, teacher rather difficult to directly provide the material contained the source book, but teachers teach them how to read and write Arabic text first. This condition is one factor inhibiting the process of learning Arabic in the program, so it takes a high level of creativity of teachers.

Teacher (*murshid*) involved in the implementation of the learning process as much as 123 people and assisted by *mushrifs* many as 22 people. The *mursyid* and *mushrifs* are lecturers of Arabic at UIN Alauddin Makassar and lecturers of Arabic from other universities who were in Makassar, alumnus various universities in Indonesia even college middle east, over the semester students who have good Arabic language skills. Special teaching second and third criteria of selection tests should follow the form of the interview, read the book of bare, written exams, and teaching practices.

Task *mursyid* and *mushrifs* are: 1) starting and ending learning process in a timely schedule, 2) do control the presence of students both in the classroom as well as the implementation of the prayers, 3) motivate students to actively follow all the activities carried out, 4) to conduct an evaluation and control sustainable manner and the achievement of student learning outcomes, and 5) maintain cooperation with program managers in addressing the problems that arose in the study.

C. Component Process Intensification Program in Foreign Languages UIN Alauddin Makassar

Data obtained associated with process components consist of data about the curriculum and teaching materials, teaching and learning, enforceability of activities on schedule, utilization of infrastructure to the maximum in order to achieve program objectives.

In the process of teaching and learning, and *mushrifs mursyid* refers to instructional materials and learning resources that have been prepared for the purpose of learning Arabic in the intensification of foreign language program targeting four keterampilan not speak but i lan more difokuskan on speaking skills alone.

Instructional materials referenced in learning designed communicative, applicable as well to summarize some form of conversation that are relevant to the learners beginners or learners who already have a basic knowledge of Arabic. It is intended that each student can understand any form of conversation contained in the instructional materials and then be able to apply it in daily life.

Associated with the learning process, berdasarkan observations that the *murshid* and *mushrifs* have different tendencies in classroom management. For teachers who have a good understanding in managing a class, tend to use models and learning techniques are varied in accordance with the characteristics of the material being taught so that the learning process takes place with very pleasant. Teachers were managing a class with an approach that varied was the lecturer of the faculty tarbiyah and teacher and lecturer of the faculty culture and humanities. Their involvement would greatly support the successful implementation of the activities in the program, even teachers can monitor the progress of knowledge and language skills of students at the faculty level, so that the activities in the program can be harmonized with the Arabic language learning activities at the faculty.

Different from the above conditions, teacher educational background of the Middle East as well as native speakers who came from Sudan tend to use methods *qawa'id tarjamah mubasyarah* and *methods*, so that the learning process looks very monotonous that resulted in the saturation of students. Despite the fact that their involvement in the program is very relevant

to the objectives that must be achieved on lessons student, para because the teachers are able to communicate in Arabic actively, as well as correcting errors common saying pronounced students, they always greet and interact with students in the Arabic language so that eventually can form exercises hear students speaking skills.

D. Component Products Intensification Program in Foreign Languages UIN Alauddin Makassar

The data obtained related to product components consist of data on changes in student learning after following the intensification of the foreign language program, with a focus observations on whether the set goals have been achieved? This means that if a student has been skillfully communicate in Arabic after following the teaching and learning process in the program ?, and obtained impacts whether the student after passing through a long process ?.

To determine whether the program objectives achieved, then at the end of the semester is done the final test. Data obtained student learning outcomes at three faculties in question can be seen below.

Table 1. Value Arabic student at the Faculty of Science and Technology

NO.	DEPARTMENT	VALUE				
		A	B	C	D	E
1.	Biology A, B	60	9	0	0	2
2.	Physics A, B	37	20	1	0	6
3.	Chemistry A, B	32	24	1	0	2
4.	Mathematics A, B, C	49	19	0	0	0
5.	City Planning A, B	17	9	0	0	21
6.	Animal Science A, B	22	48	4	0	4
7.	Information Systems	27	0	0	0	2
8.	Architecture A, B, C, D	43	39	1	0	15
9.	Informatics A, B, C	48	27	6	0	12

From the table above it is understood that there is a 54.82% who received grades A, 32.73% who earn a grade of B, 1.96% who earn a grade of C, 0% were scored D, and 10.47% are scored E. the data shows that learning Arabic at the intensification of foreign language program in the faculty of science and technology included in the category satisfactory, considering the student in the faculty mostly from public schools who have never learned Arabic before.

Table 2. Value Arabic students of Faculty of MT and Teaching

NO.	DEPARTMENT	VALUE				
		A	B	C	D	E
1.	Islamic Education 1,2,3,4,5,6,7,8	119	33	0	3	4
2.	Islamic Education Management 1,2,3,4	22	43	0	0	3
3.	Teacher Education Madrasah Ibtidaiyyah	22	12	2	0	1
4.	Arabic Education 1.2	40	0	0	0	0
5.	English Education 1,2,3,4	39	39	3	0	6
6.	Biology Education 1.2	10	22	2	0	4
7.	1,2,3,4 Physics Education	41	22	4	0	5
8.	Mathematics education	36	2	0	0	0

The above data show that there are 61.04% who received grades A, 32.10% were obtained value B, which scored 2.04% C, 0.55% who got a D, and 4.27% are scored E. From these data it can be concluded that learning Arabic in the intensification program in a foreign language and teacher training faculty tarbiyah very satisfying.

Table 3. Value Arabic student at the Faculty of Sharia and Law

NO.	DEPARTMENT	VALUE				
		A	B	C	D	E
1.	Criminal law and constitutional 1,2,3,4,5,6,7,8	93	74	15	0	23
2.	Religious Courts 1,2,3,4,5,6,7,8,9,10	204	34	2	0	5
3.	Legal studies 1,2,3,4,5,6	123	22	0	0	6
4.	Legal Studies 7,8,9,10,11,12,13,14	72	79	11	4	30
5.	Comparison of the School and the Law	26	17	0	0	5

The above data show that there are 61.30% who received grades A, 26.74% were obtained value B, which scored 3.31% C, 0.47% who got a D, and 8.16% are scored E. From these data it

can be concluded that learning Arabic at the intensification of foreign language program at the faculty of sharia and the laws that are in the satisfactory category.

Of the three faculties of the sampled tend to show similar results. This indicates that the general intensification of foreign languages at UIN Alauddin Makassar runs as expected. Although there are some problems in the learning process, but these barriers can be overcome by the teachers and managers.

IV. CONCLUSIONS AND RECOMMENDATIONS

Based on the description of the results of research and discussion, the conclusions of a special that can be drawn from this study are: 1) the suitability of intensification of foreign languages for new students viewed from component *context* is appropriate, 2) the condition of the factors supporting the intensification program of foreign language seen from *the input* component is already well, 3) the implementation of the program of foreign language intensification of *process* components running properly, 4) foreign language program seen intensification of already successful *product* components.

However there are some things that must be considered by the teachers and managers for programs to run with better and output more quality, among other things: 1) the need to do a placement test and students are grouped according to the results of these tests, 2) need to develop teaching materials kontekstual based on the objective conditions of students in each group, 3) the need to provide the means prasana sufficient learning, 4) there needs to be a real effort in improving teacher training, and 5) program providers need to establish good cooperation with the leadership, hand majors as well as various student organizations observer and lover of the Arabic language to intensify the acquisition and mastery of the Arabic language students.

BIBLIOGRAPHY

- [1] Creswell, John W. Educational Research, Planning, Conducting, and Evaluating Quantitative and Qualitative Research. University of Nebraska – Lincoln, Ohio, 2008.
- [2] Gay, L.R. Educational Research, kompetencies for Analisis and Aplication, Second Edition, Florida International University.
- [3] Guba, Egon G and Yvonna S. Lincoln. Effective Evaluation: Improving the Evaluation Results Through Responsive and Naturalistic Approaches, San Francisco, Jossey, Bass Publisher, 1981.
- [4] Mc David, James C., Laura R.L.Hawthorn, Program Evaluation & Performance Measurement, London, Sage Publications, 2006
- [5] Patton, Michael Quinn, Qualitative Evaluation and Research Methods Newbury Park, California, Sage Publication, 1990.
- [6] Brinkerhoff, Robert, Program Evaluation, Boston, kluwer Nyhoff Publishing, 1986.
- [7] Brinkerhoff, Robert, Brethower, Dale M., Hluchy, Terry Nowakowski, Jerry Ridings, Program Evaluation a Practitioner's Guide for Trainer & Educator, Boston, Kluwer-Nijhoff Publishing, 1983.
- [8] Chen, Huey-Tsyh, Practical Program Evaluation. United States of America, Sage Publications, 2005.
- [9] Davidson, E. Jane. Evaluation Methodology Basic, United States of America, Sage Publication, 2005.
- [10] Stufflebeam, Daniel L & Anthony J. Shinkfield, Evaluation Theory, Models, and Application, USA, John Willey & Sons, Inc. 2007.
- [11] Royse, David., Thyer, Bruce A.m Padgett, Deborah. K., Logan, Program Evaluation an Introduction, Fourth Edition, Belmont USA, Thompson Brooks/Cole, 2006.

EVALUATION OF THE CIVILIZING MORAL CHARACTER IMPLEMENTATION IN ELEMENTARY SCHOOL

Sulthoni

Universitas Negeri Malang, Jl. Semarang 5 Malang
tonikueren@gmail.com

Abstract - The purpose of this research was to describe the evaluation of civilizing moral character in elementary school environment by using qualitative descriptive method. The data collection is by interview, observation, documentation. Subject of research is in a public elementary school. The results of this study indicate that there are some important factors that fostered, including: faith, devotion, honesty, ideals, democratic atmosphere, caring, openness, unity, security, hygiene, health, beauty, and manners. The implementation of civilizing moral character in the school environment has been created properly.

Keywords: *Evaluation, civilizing, moral character*

I. INTRODUCTION

Mental health, or the noble character noble spirit is very important for the development of civilization and culture of a nation, in addition to intelligence thought and intellectual capabilities. Efforts to improve the intelligence of thinking, mental development, manners and noble character is a shared task between families, schools and communities.

Thus it can be understood that the behavior and immoral is caused by the low moral. Moral is low due to the moral character education less effective. Santoso (1991: 27) revealed that the affairs of depravity can not be repaired only by an appeal, speeches, sermons, plays, seminars, working meetings, and various forms of other similar efforts, but should be with the efficiency of moral character education in families, schools and communities.

Planting character value needs to be balanced with the state of a supportive environment. Character education is not just a school assignment, but the task of all of us, even the family, school, community and government. Togetherness that can be produced from the fruit of moral character education.

The school curriculum starting from the most basic to the highest level allocate ample time for the field of study of potential for moral character guidance, among other Religious Education, Citizenship Education (Civics), and Social Sciences. But it does not mean that the moral character education in schools has been successfully held with satisfactorily, because the learning process is carried out by the approach of memorization (cognitive) are oriented solely to get good grades, then how subjects can impact the behavior change, was never considered. What happens is that the gap between the moral character knowledge (cognitive) and behavioral (action) (Megawangi, 2004: 80).

Character education is taught in schools for the purpose, among others, to build a future generation so besides being smart also have a certain noble and virtuous character, in accordance with the objectives listed in the National Education Act of the Republic of Indonesia No. 20 of 2003, Chapter II, Article 3 formulate that: The purpose of National Education is for developing students' potentials to become a man of faith and devoted to God Almighty, noble, healthy, knowledgeable, skilled, creative, independent, and become citizens of a democratic and responsible.

We as the future generation partly been affected by things that are negative that leads to deviant behavior and poor manners (Rusyan, 2003: 16).

In responding to misbehavior, the government issued a regulation the Minister of National Education No. 21 of 2006 dated May 23, 2006 on Standards of Competency Education Unit. Education Graduate Competency Standards for SD/MI/SDLB / Package A includes: (1) Running professed religious teachings according to the stages of child development; (1) Running professed religious teachings according to the stages of child development; (2) Know the advantages and disadvantages of self; (3) Comply with the social rules that apply in the environment; (4) Respecting the diversity of religion, culture, ethnicity, race, and socioeconomic groups in the surrounding environment; (5) Using the information about the environment around logically, critically and creatively; (6) Demonstrated ability to think logically, critically and creatively with the guidance of a teacher / educator; (7) Indicates a high curiosity and realize their potential; (8) Demonstrated ability to solve simple problems in everyday life; (9) Demonstrated ability to recognize the natural and social phenomena in the surrounding environment; (10) Showing love and concern; (11) Shows the love and pride to the nation, the state and the homeland of Indonesi; (12) Demonstrated ability to perform activities of local arts and culture; (13) Indicates clean living habits healthy, fit, secure, and take advantage of free time; (14) communicating clearly and politely; (15) In cooperation with a group, mutual help, and keep yourself in a family environment and peers; 16) Showing indulgence to read and write; (17) Indicates listening skills, speaking, reading, writing, and arithmetic.

II. RESEARCH METHODS

To assess the civilizing character in elementary school principal elements which correspond to the formulation of the problem, objectives, this research uses descriptive method with qualitative approach.

A. Data Source

Sources of data in this study include: primary data sources and secondary schools in the acculturation of character for learners obtained from the principal and teachers, guards and cleaners Elementary School in Malang

B. Data Collection Technique

In gathering the data researchers applied the technique of deep observation, interviews, documentary studies and active participation in the activities of teachers, principals, school devices. Data collection techniques as follows: 1) observation, 2) Interviews, and 3) Studies Documentation.

C. Data Analysis Technique

Qualitative data analysis performed interactively and continues at every stage of research so as to completion and until saturated. data analysis steps are as follows: data reduction, data presentation and conclusion, verification.

D. Testing Data.

Testing the credibility of the data is done by: extension, observation, improve endurance, triangulation, inspection peers, negative case analysis, member check (checking member)

E. Research Subject

The subject of research as follows: Principal teachers, administration, security guards, and the guards cleanliness Elementary School.

III. RESEARCH RESULT

The success of creating a school environment that is conducive to the civilizing character there are some important factors that need to be cultivated, namely:

A. Faith

Planting of faith in the students a lot of work involved, among other things: to celebrate the Islamic holidays, Pondok Ramadan in the month of Ramadan, zakat, slaughtering the sacrificial animals on the feast of sacrifice. With the existence of these activities show that the school hopes to add to the faith on the learner.

B. Devotion

Activities undertaken school is midday prayers in congregation between teachers and students in higher classes and take turns at the mosque school. Islamic lectures held once every two days were held in the morning before going to school is housed in a nearby mosque,

implementation and day turns each class. One day for grades 1-3 and the second day for grades 4-6. In the month of Ramadan Ramadan activities held cottage Duha prayer together, and lectures between teachers and learners in a nearby mosque.

C. Honesty

Teachers familiarize learners to honesty, for example, when teachers held a replay, to correct it submitted to the learners, here learners trained honesty in correcting and giving the score.

D. Modeling

Teachers every morning standing in front of the school entrance welcoming incoming students with a handshake. Security guard set the path and help get across learners. Guard hygiene cleaning the school yard. This indicates that the start of the guard of cleanliness until the principal became role models for students.

E. Democratic atmosphere

At the meeting time teachers, teachers of freedom of speech. So also between teachers and learners, learners' freedom of speech and mutual respect. This shows that the democratic atmosphere has been developed.

F. Concern

Every Friday, students contribute voluntarily and donated by, among others, contributed to the students who was ill at the time of teachers and friends dwindled. This shows their concern for fellow teachers, teachers with students, and students with student.

G. Openness

At the time there was a meeting between the school and parents. The school principal and the school committee report school programs and school finance. This suggests that openness has been developed at the school.

H. Togetherness

Each class formed the community of parents. Its activities include: a study tour. In a study tour intertwined relationship, communication, mutual respect, mutual help among teachers, parents of students, and students. This suggests the establishment of togetherness.

I. Security

Teachers and students feel peaceful, peaceful, and safe in the school environment, Because teachers live in harmony, there is no dispute between the teachers and the security environment well preserved, that in addition to the security guard in the school environment there is also the official residence occupied by teachers who do not have a home and at the same time the teachers care for the school environment. This suggests that a safe atmosphere be maintained.

J. Order

Teachers in implementing the learning in accordance with the specified time. There is a third grade teacher who called in sick, schools have set up temporary replacement teachers to teach, so that the learning process is still running. This shows that the order in the school environment have been implemented.

K. Cleanliness

Cleanliness guard every morning cleaning the school yard, and in each classroom was cleaned by students picket. Every corner there is a trash can, so the atmosphere in the school environment look clean. If students do not throw garbage in its place, a guard or teacher who is viewed advised. This shows that the cleanliness in the school environment goes well.

L. Health

Residents school every Friday morning held a joint exercise in the morning, the principal, teachers, and students. In the classroom every completed hour picket the education of students who cleans each class, so tomorrow morning ready for learning. So also the cleanliness of the psychic that the spiritual cleansing is held once a week from grade 1 to grade 6 turns. It is intended to raise awareness of psychic hygiene. The activity shows the attempt to maintain both physical and psychological health of the citizens of the school

M. Beauty

Trees and flowers in the school yard has been neatly arranged. So it looks cool and shady. Well-organized teachers' room, each teacher has their desks, so that teachers before teaching, after teaching, or during hours of rest can work related to teaching. Principal's room neatly arranged and at the same time as the living room.

N. Manners

Citizens schools as a model in implementing the manners of students. Attitudes, behavior and speech-mannered is a must for people in schools, ranging from janitors to the principal. In this school has applied it. If there are students who do deviate from the rules governing the manners, the teacher immediately reprimanded and advised.

The findings indicate that in the school environment has been created familiarization character.

IV. DISCUSSION

Minister of Education and Culture has inaugurated Anies Baswedan character Growth program (PBD) in Jakarta, Friday (24/7). The program is a positive attitude and behavior habituation applied to students since the orientation of new students until graduation. The purpose of implementation of character growth movement is making the school as a garden to cultivate positive characters of students at all school levels. There are six stages of program implementation flow growth of manners, namely the stage of teaching, habituation, training to be consistent, the process of habituation, the formation of character, and a culture.

It shows that the government is serious about dealing with character education, given the reigns now willingly acts that lead to immoral as sex abuse, rape and mass murder. Thus the need to start early age receive moral character education. At school there needs to be a civilizing manners at all levels in order to prevent the things that lead to immoral. In schools that were subjected to this research at least has been carrying out a civilizing character by applying the 14 components that need to be developed, namely: (1) Faith. Faith greatly affect a person's behavior. That faith needs to be fostered and cultivated according respective religious beliefs. In the school to instill faith in a lot of work involved, the faith is expected that each student can build itself into a human, noble character. (2) devotion. Devotion should be inculcated early to learners enter school through various activities, Such activities have reflected piety, so that the human being virtuous noble character. Because basically the human quality is determined by devotion. devotion is a reflection of the values of faith in the form of behavior which manifested itself in running away from the religious orders and prohibitions. Piety is a reflection of the values of faith in the form of behavior which manifested itself in running away from the religious orders and prohibitions. (3) Honesty. In many ways, attitudes and behavior do not lie, do not cheat, brave and willing to sacrifice for the sake of truth and admit mistakes. These actions should be realized and cultivated so that it becomes part of the everyday life, whether related to God, oneself and with others. This is consistent with the results of the research. Thus honesty becomes assertive attitudes and behaviors that must be implemented. (4) Modeling. Modeling is a key character in the acculturation. Starting from maintaining the cleanliness until the principal became role models for students. Modeling is much more important than giving verbal lessons, because the example is giving an example by deed or action. This school has implemented well. (5) democratic atmosphere. Democratic atmosphere in question is to respect the rights of others to express opinions, advice, expression, creativity. The atmosphere in school should be an atmosphere that shows their freedom of speech, and respect differences of opinion in accordance with democratic manners. This shows that the democratic atmosphere has been developed at the school. (6) Concern. Concern manifested among others, in empathy and edify each other, tell each other, remind each other, love each other and protect each other so that any problems can be solved more quickly and more easily. Refraction themselves have a concern in the school environment needs to start early. It has been implemented in schools. (7) Openness. School management system should be transparent, meaning that every activity should be carried out openly, particularly pleasing with financial problems and in making decisions. Open management will eliminate mutual suspicion, prejudice, and eliminate defamation. This open management has been practiced by the principal, employee administration, teachers, and by the students. This suggests that openness has been developed at the school. (8) Togetherness is the atmosphere of community relations in the school system are reflected in the attitudes and behaviors such as helping, tolerance, mutual respect, and open. Togetherness is aimed to strengthen the relationships between principals, teachers, students and other school communities to realize an

atmosphere of brotherhood in grammar school harmonious relationship. It has been implemented in schools that show the intertwining of togetherness. (9) Security. Security here is intended as a sense of security and peace, free from fear, both physically and mentally. Security is the principal capital to create a harmonious school atmosphere and fun. It is evident that teachers and learners to feel peaceful, peaceful, and safe in the school environment, this case shows that the atmosphere is safe well maintained. (10) Order. Orderliness is a condition that reflects the harmony and order in the interaction between the school community. Order in the school environment has been going well. 11) Cleanliness is part of faith. The ambiance is clean, tidy and refreshing will give the impression of fun for the citizens of the school. Schools have grown psychic hygiene. 12) Health. Health also involves physical and psychological aspects. Physical health for people in schools let pursued by way of exercising regularly, eating nutritious food. While mental health is built in a way let incited. It shows the attempt to maintain the health of the school community. 13) Beauty. The beauty here is meant the atmosphere of good school environment office, the teacher's room, library and classrooms It is evident that the school community as a role model in implementing manners learners. Attitudes, behavior and speech-mannered is a must for people in schools, ranging from janitors to the principal. In this school has applied it.

V. CONCLUSION

Schools as formal education, its role inculcate moral character education begins from the principal to the messenger. Togetherness is creating a school atmosphere conducive for the cultivation of character to make the school better provide meaningful learning experiences for students.

Implementation of familiarization character in the target schools have carried out 14 components properly. although there are still some that need to be improved, both physically and activities that lead to activities manners. This shows that the school environment has created a civilizing moral characterity.

REFERENCE

- [1] Balitbang-Puskur. (2001). *Kurikulum Berbasis Kopetensi Mata Pelajaran Budi Pekerti* untuk Kelas I-VI, Buran ke 6. Jakarta: Depdiknas.
- [2] Departemen Pendidikan Nasional. (2004). *Pedoman Penciptaan Suasana Sekolah yang Konduktif dalam Rangka Pembudayaan Budi Pekerti Luhur Bagi Warga Sekolah*. Buku II. Jakarta: Departemen Pendidikan Nasional.
- [3] Megawangi, R. (2004). *Pendidikan Karakter*. Jakarta: Indonesia Heritage Foundation.
- [4] Miles, M. B. dan Huberman A.M. (1992). *Analisis Data Kualitatif*. Terjemahan. Jakarta: UI Press.
- [5] Moleong, L.J. (2007). *Metodologi Penelitian Kualitatif*. Bandung: PT. Remaja Rosdakarya.
- [6] Nasution. (1992). *Metode Penelitian Naturalistik Kualitatif*. Bandung: Tarsito.
- [7] Rusyan, A. T. (2003) *Pendidikan Budi Pekerti*. Jakarta: PT. Intimedia Ciptanusantara. Santoso, R.S.I. (1991). *Batu Landasan Ketertiban Masyarakat: Pembinaan Watak yang Kokoh*. Mimbar Pendidikan. 1(X): 26 – 27.
- [8] UUSPN. (2003). *Undang-Undang Sistem Pendidikan Nasional*. Jakarta: Sinar Grafika.

THE EVALUATION OF 2013 CURRICULUM IMPLEMENTATION ON THEMATIC INTEGRATIVE TOWARD MATH SUBJECT FOR ELEMENTARY SCHOOL IN EAST LOMBOK

Syukrul Hamdi
STKIP Hamzanwadi Selong,
e-mail: syukrulhamdi@gmail.com

Abstract - This research is aimed at describing the effectiveness of thematic integrative learning on math in East Lombok in relation to the implementation of 2013 curriculum. The evaluation approach used is CIPO model. The research population consisted of all elementary schools in East Lombok Regency adopting the 2013 Curriculum, with a total of 30 schools. Of these 30 schools, 6 pilot schools and 2 independent schools were selected randomly to be used as the research sample. Each school will be represented by School Headmaster, 4 School Teachers, 30 students of grade 4 and 2 of Elementary school's supervisors. Data will be collected using questioner as the main instrument and mathematic test to know the students' achievement on thematic integrative of math content. The result shows that 1). the implementation of the 2013 Curriculum is effective in terms of the context for the regency town, district town and rural areas; 2). the implementation of the 2013 Curriculum is deemed ineffective for rural elementary schools, less effective for elementary schools in the district town and effective for elementary schools in the regency town; 3) the implementation of the 2013 Curriculum is effective in terms of the process for the regency town and district town and less effective for the rural area and, finally 4) the implementation of the 2013 Curriculum is effective in terms of the output for the regency town and less effective for the district town and the rural area.

Keywords : 2013 Curriculum implementation, Mathematic, CIPO model.

I. INTRODUCTION

Curriculum implementation in the Indonesian education focuses on two types of curriculums, namely KTSP (the Educational Unit Level Curriculum) and the 2013 Curriculum. The use of both curriculums is adjusted to the conditions of each school, especially in terms of preparedness of the facilities and infrastructure which support such implementation. Thus, each region prefer to map schools that are already prepared and those that are not therefore the educational objectives can be optimally achieved.

Differences in terms of the competencies contained in each curriculum will definitely affect the concept or materials which the students master. Elementary school students will acquire a level of basic perception and knowledge which is different from that of junior and senior high school students. The materials included in KTSP remain specific and not integrated with the other subjects, while in the 2013 Curriculum, the nature of the materials is thematic-integrated, grounded in the theory of development of students' perception for a number of subjects, including mathematics.

Mathematics as part of the disciplines studied in every level of education contains interrelated concepts. In order to master the next level, students are required to study the previous level. One who wishes to be a mathematician must first learn it from the very beginning (Erman Suherman et al., 2001:59). Therefore, it is imperative to deliver the strong instruction of the essentials of mathematics to all students, especially the elementary school students.

Basically, differences in materials in the existing educational levels will have an impact on the students' level of mastery, moreover the curriculums adopted are up-to-date or the latest, especially for the 2013 Curriculum. In the 2013 Curriculum, the classroom instruction is more

student-oriented and teachers merely function as a facilitator and mediator guiding and directing these students. In other words, the students are directed and stimulated simultaneously to figure out their problems and solutions to those problems on their own. Such conditions may pose an obstacle, especially for schools which are still not ready to implement it, both in terms of human resources, particularly the teachers, and supporting facilities, especially the necessary textbooks, as well as the role of parents.

Description of the issues on the implementation of the 2013 Curriculum 2013, particularly in relation to mathematics instruction, are also found in East Lombok Regency. These are indicated by the results of observations and interviews carried out in nine elementary schools situated in this region. Almost all the principals and teachers deal with the same problems, namely the teachers' inadequate understanding of the real implementation of the curriculum because only a limited number of educators or teachers can attend training, resulting in teachers' lack of mastery of the evaluation or assessment system which shall be carried out at each instructional process. In addition, on the average, the schools implementing the 2013 Curriculum of the independent category designated to apply it are not equipped with facilities in terms of the procurement of textbooks and handbooks, therefore a book that should be used for one meeting is often used for two or three meetings instead. On the other hand, the 2013 Curriculum requires that the materials should be up-to-date or adjusted to the latest issues.

The problems concerning the implementation of the 2013 Curriculum, particularly in relation to mathematics, are not only found in schools in rural/ mountainous areas, but also in the urban area of East Lombok Regency. Hence, the effectiveness of the 2013 Curriculum implementation in relation to the thematic content in the subject of mathematics needs evaluation by conducting scientific research. Evaluation is defined as the process of providing information, in which the information provided must be reliable, effective, cost-efficient, efficient, safe, useful, and trustworthy (Stufflebeam and Shinkfield, 2007:4). Based on the foregoing, the author is interested in conducting research entitled "The Evaluation Of 2013 Curriculum Implementation On Thematic Integrative Toward Math Subject For Elementary School In East Lombok".

II. RESEARCH METHOD

This is evaluation research employing the CIPO model focusing on the aspects of context, input, process and output. It employed qualitative and quantitative approaches. The research population consisted of all elementary schools in East Lombok Regency adopting the 2013 Curriculum, with a total of 30 schools, 9 of which are pilot schools and the remaining 21 schools are independent schools. Of these 30 schools, 6 pilot schools and 2 independent schools were selected randomly to be used as the research sample. This sample were divided into three areas, namely the regency town, the district town and rural/ mountainous areas. The sample schools were SDN 3 Selong, SDN 3 Pancor, SDN 3 Rempung, SDN 1 Anjani, SDN 3 Aikmel, SDN 7 Lendang Nangka, SDN 3 Masbagik Timur and SDN 7 Danger.

Each sample school consisted of the principal, School Headmaster, the fourth-grade teacher, 30 fourth-grade students and two elementary school supervisors as the data sources. The data were gathered using questionnaires as the main instrument and mathematics tests to determine students' learning outcomes in relation to the thematic-integrated content in the subject of mathematics. The instrument validity was ensured by the content validity involving seven experts, namely 1 psychologist, 2 measurement experts and 4 experts in mathematics and mathematics instruction. The results were analyzed using the Aiken Formula.

In relation to the qualitative data analysis, the quantitative data obtained through the assessment instrument were converted to qualitative data using a scale of 4 and then described. The resulting descriptions were used as the standards for the quality of the input, process and output aspects of the elementary schools in East Lombok Regency. The conversion tables are presented in Tables 1, 2 and 3 below.

Table 1. Quantitative to Qualitative Data Conversion for the Input Aspect

Teacher		Student	
Score	Category	Score	Category
$X > 55.25$	Very Effective	$X > 6.5$	Very Effective
$42.5 < X < 55.25$	Effective	$5 < X < 6.5$	Effective
$29.75 < X < 42.5$	Less Effective	$3.5 < X < 5$	Less Effective
$X < 29.75$	Ineffective	$X < 3.5$	Ineffective

Table 2. Quantitative to Qualitative Data Conversion for the Process Aspect

Teacher		Student	
Score	Category	Score	Category
$X > 32.5$	Very Effective	$X > 22.75$	Very Effective
$25 < X < 32.5$	Effective	$17.5 < X < 22.75$	Effective
$17.5 < X < 25$	Less Effective	$12.25 < X < 17.5$	Less Effective
$X < 17.5$	Ineffective	$X < 12.25$	Ineffective

Table 3. Quantitative to Qualitative Data Conversion for the Output Aspect

Score	Category
$X > 75.25$	Very Good/Very Effective
$50.5 < X < 75.25$	Good/Effective
$25.75 < X < 50.5$	Less Good/Less Effective
$X < 25.75$	Poor/Ineffective

III. RESEARCH FINDINGS AND DISCUSSIONS

A. CONTEXT EVALUATION

According to Kaufman & Thomas (Mukhadis, 2013: 118), in program evaluation activities, a context refers to a matter of external existence to the program itself and it is a condition which has already existed in a program (the antecedent variable), however it will be able to influence the process of implementing and realizing the predetermined objectives of a program.

The component of contexts plays a vital role in the CIPO model because it is directly related to the input, process and product as suggested by Scheerens (2004) that the dimension of contexts serves as a source of inputs and as a generator of output simultaneously. Furthermore, Scheerens (2004:17) states that a context's indicators relate to the standards and systems of the national education, including the component of the context of the school environment, such as the demography, the economic and financial conditions, the educational objectives and standards, and the fact whether the school building is compatible with the education standards or not. The component of contexts refers to any matters related to the specific environment of an educational institution or an educational program.

Based on the results of the observations and documentation in 8 sample schools, it was revealed that schools in the rural areas, the district town, and the regency town did not have any significant obstacles. In general, the recapitulation of context evaluation showed effective results, meaning that public elementary schools (in Indonesia, these schools are known as Sekolah Dasar Negeri which are often abbreviated as SDN) in the rural areas, the district town, and the regency town in East Lombok support the implementation of the 2013 Curriculum. It also indicates that the schools' vision, mission, environment, and programs also support the implementation of the 2013 Curriculum in the area of East Lombok Regency. Nevertheless, there were a number of aspects deemed ineffective, namely the understanding of several teachers regarding the implementation of the 2013 Curriculum and the economic and financial conditions, especially in schools implementing the 2013 Curriculum independently.

B. INPUT EVALUATION

According to Scheerens (2004:17), inputs include both financial and resources invested in education, such as the costs for each students, for research and development in education, the percentage of workers working in the field of education the ratio of teachers to students in each educational level as well as human resource characteristics, age, gender, experience, qualifications and salaries of teachers. Therefore, in terms of the input, the evaluation was focused on the teachers and students. In relation to the teachers, the evaluation highlighted the lesson planning, learning experiences and preparedness of the learning facilities and infrastructure. In relation to the students, the evaluation focused on the role of parents and the students' preparedness related to the implementation of the 2013 Curriculum. The evaluation results are shown in Table 4.

Table 4. The evaluation results for the Input Aspect

Location of Schools	Score	Category	Score	Category
Elementary Schools In Rural Areas	50.00	Effective	3,00	Ineffective
Elementary Schools In District Town	52.67	Effective	4,67	Less Effective
Elementary Schools In Regency Town	58.00	Very effective	6,00	Effective

The implementation of the 2013 Curriculum is effective based on the result of the input evaluation for public elementary schools in the urban areas at the regency level and at the subdistrict level viewed from the component of teachers. On the contrary, such implementation is ineffective in terms of the inputs for elementary schools in rural areas, while for elementary schools in the urban areas at both the regency and subdistrict levels, it is less effective. Public elementary schools in the urban areas at the regency level did not encounter problems related to the aspect of inputs because in terms of the environment, especially support from parents and the community are positive as indicated by their maximum efforts to meet the instructional demands, ranging from providing learning materials to purchasing books, stationery and other supporting facilities. A quite different finding was found in rural areas and some urban areas at the subdistrict level as some parents and the community do not really know and care about their children's learning needs. This is because they give priority to work and food as the majority of the guardians or parents of the students are farmers, farm laborers, stone crushers, motorcycle-taxi driver, sandpaper men, and the like. The general description of schools in the rural areas and the urban area at both the regency and subdistrict level is presented in Figure 1.



Figure 1. The general description of schools

The other causes of less effective implementation of the 2013 Curriculum in relation to mathematics in rural areas are due to the aspects of school management, teachers' competencies as well as facilities and infrastructure of public elementary schools in the rural area of East Lombok Regency which do not support such implementation. This is evident from the inadequate number of teachers participating training in the implementation of the 2013 Curriculum, especially schools of the independent category; the limited number of additional books and learning materials, especially the books available in the library, the unavailability of the internet and technology-based instructional media which are vital for updating the teaching materials.

C. PROCESS EVALUATION

According to Stufflebeam and Shinkfield (1986:341) the component of processes includes identifying, predicting weaknesses in terms of the procedural design or the implementation,

informing the program decision, recording, procedures and the activities. In relation to the evaluation process on both the teachers and students of elementary schools in East Lombok Regency, the implementation of the 2013 Curriculum is generally effective, except for the rural elementary school students who belongs to the less effective category. This suggests that the aspects of lesson planning, the implementation and the assessment support the implementation of the 2013 Curriculum. The questionnaire results are presented in Table 5 below.

Table 5. The evaluation results for the Process Aspect

Location of Schools	Score	Category	Score	Category
Elementary Schools In Rural Areas	30	Effective	16.00	Less Effective
Elementary Schools In District Town	32	Effective	18.67	Effective
Elementary Schools In Regency Town	32	Effective	21.00	Effective

Some ineffective aspects of the process variable, especially for rural areas, are lesson planning and its implementation. Some educators or teachers have attempted to use the standards of processes during the instruction, however it is not supported by necessary instructional media. Moreover, the the standards of processes which have been established are difficult to implement due to lack of support from the supporting facilities and infrastructure, such as teacher handbooks and student worksheets necessary to explore and improve students' competencies which eventually will lead to an assessment which remains focused on tests rather than performance (process-related assessment/ authentic) and less up-to-date information, especially in accordance with the content of the materials to be delivered.

D. OUTPUT EVALUATION

The main focus of the output evaluation in this research is the students' learning outcomes. The output evaluation is used as a basis for determining the resulting learning outcomes and improving the instructional process. In practice, this output evaluation is always associated with the extent to which an educational institution manages to achieve its objectives (Cuyvers, 2002:122). The successful achievement of the goals of education, both at the national and institutional levels as well as the level of mastery of the competency standard/ basic competency of a subject, lies on the students. To determine the students' learning outcomes, this conducted tests given in two stages.

Viewed from the output aspect, the implementation of the 2013 Curriculum is effective for the regency town, less effective for the district town and ineffective for the rural area. This suggests that academic achievement in the form of the knowledge aspect demonstrated by the students' learning outcomes in mathematics, especially for elementary schools in East Lombok Regency, is higher in the regency town than those in the district town and the rural area. The results can be seen in Table 6 below:

Table 6. The evaluation results for the Output Aspect

Location of Schools	Score Tes 1	Category	Score Tes 2	Category
Elementary Schools In Rural Areas	24.74	Poor	28.96	Less Good
Elementary Schools In District Town	34.08	Less Good	38.00	Less Good
Elementary Schools In Regency Town	53.60	Good	52.86	Good

IV. CONCLUSIONS AND SUGGESTIONS

The research conclusions are drawn as follows: 1). the implementation of the 2013 Curriculum is effective in terms of the context for the regency town, district town and rural areas; 2). the implementation of the 2013 Curriculum is effective in terms of the input in relation to the elementary school teachers for the regency town, district town and rural areas. As for the input in terms of the students, such implementation is deemed ineffective for rural elementary schools, less effective for elementary schools in the district town and effective for elementary schools in the regency town; 3) the implementation of the 2013 Curriculum is effective in terms of the process for elementary schools in the regency town, district town and rural areas both in relation to the teachers and the students, except for the process aspect of rural elementary school students which is belong to the "less effective" category; and, finally 4) the implementation of the 2013 Curriculum is effective in terms of the output for the regency town and less effective for the district town and the rural area.

Based on the evaluation results, it is suggested that generally the aspects of the context and the process neither pose problems nor inhibit the implementation of the 2013 Curriculum in elementary schools in East Lombok Regency, it indicates that most teachers have already

understood and can implement the 2013 Curriculum properly. It is suggested that the government and policy makers should give special attention to the input aspect, especially in terms of facilities, and help the parents to understand the instructional activities carried out in the implementation of the 2013 Curriculum, especially in rural elementary schools, through various programs involving participation of parents, school components and the community in the hope that there will be no gap between the regency town, the district town and the rural area in terms of inputs and outputs.

REFERENCES

- [1] Cuyvers, G., (2002) *Kwaliteitsontwikkeling in het Onderwijs*. Apeldoorn: Garant Uitgeverij N.V
- [2] Erman Suherman, et.al (2001). *Strategi pembelajaran matematika kontemporer*. Bandung: JICA UPI
- [3] Mukhadis.A.(2013). *Evaluasi program pembelajaran bidang teknologi*. Malang: Bayumedia Publishing.
- [4] Scheerens, J., (2004) *The quality of education at the beginning of the 21st Century*, (Paper for the UNESCO Education for All Global Monitoring Report) data access, April, 7th 2016. <http://unesdoc.unesco.org/images/0014/001466/146697e.pdf>
- [5] Stufflebeam, D.L. and Shinkfield, A.J., (1986). *Systematic evaluation: a self-instructional guide to theory and practice*. Boston: kluwer-Nijhoff Publishing.
- [6] Stufflebeam, D.L. and Shinkfield, A.J., (2007). *Evaluation theory, models and application*: San Fransisco: Jossey-Bass